Contrasting environmental preferences of photosynthetic and non-photosynthetic soil cyanobacteria across the globe

Concha Cano-Díaz*[1], Fernando T. Maestre[2], David J. Eldridge[3], Brajesh K. Singh[4,5], Richard D. Bardgett[6], Noah Fierer[7,8], Manuel Delgado-Baquerizo[2,8]

[1]Departamento de Biología, Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología. Universidad Rey Juan Carlos, Móstoles, 28933, Spain

[2]Departamento de Ecología and Instituto Multidisciplinar para el Estudio del Medio "Ramón Margalef", Universidad de Alicante, Carretera de San Vicente del Raspeig s/n, 03690 San Vicente del Raspeig, Alicante, Spain

[3]Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences. University of New South Wales, Sydney, New South Wales 2052, Australia

[4]Global Centre for Land Based Innovation. University of Western Sydney, Penrith, 2751, New South Wales, Australia

[5]Hawkesbury Institute for the Environment, University of Western Sydney, Penrith, NSW, 2751, Australia

[6]School of Earth and Environmental Sciences, Michael Smith Building. The University of Manchester, Manchester, M13 9PT, UK

[7]Department of Ecology and Evolutionary Biology. University of Colorado, Boulder, CO 80309, USA

[8]Cooperative Institute for Research in Environmental Sciences. University of Colorado, Boulder, CO 80309, USA

*Correspondence e-mail: conchacanodiaz@gmail.com

ACKNOWLEDGEMENTS

We would like to thank Victoria Ochoa and Beatriz Gozalo for their help with soil analyses and Hugo Saiz for his help with network analyses. We are grateful to Christophe V.W. Seppey and the other two anonymous reviewers for their insightful comments and suggestions. M.D-B. acknowledges support from the Marie Sklodowska-Curie Actions of the Horizon 2020

Framework Programme H2020-MSCA-IF-2016 under REA grant agreement n°702057. The work of C.C-D and F.T.M. and the global drylands database were supported by the European Research Council (ERC Grant Agreements 242658 [BIOCOM] and 647038 [BIODESERT]) and by the Spanish Ministry of Economy and Competitiveness (BIOMOD project, ref. CGL2013-44661-R). F.T.M. acknowledges support from Generalitat Valenciana (BIOMORES project, ref. CIDEGENT/2018/041). B.K.S research on biodiversity is supported by the Australian Research Council (DP170104634). R.D.B. was supported by the UK Department of Environment, Food and Rural Affairs (DEFRA) project number BD5003 and a BBSRC International Exchange Grant (BB/L026406/1).

BIOSKETCH

Concha Cano-Díaz is a biologist interested in the distribution and ecological drivers of soil cyanobacteria. She is currently studying the effects of climate change and soil formation processes on cyanobacterial communities around the world.

DR. CONCHA CANO-DÍAZ (Orcid ID : 0000-0001-6948-6553)

PROF. DAVID ELDRIDGE (Orcid ID : 0000-0002-2191-486X)

Article type : Research Papers

Contrasting environmental preferences of photosynthetic and non-photosynthetic soil cyanobacteria across the globe

Running title: Global preferences of soil cyanobacteria

Abstract

Aim: Cyanobacteria have shaped the history of life on Earth, and continue to play important roles as carbon and nitrogen fixers in terrestrial ecosystems. However, their global distribution and ecological preferences remain poorly understood, particularly for two recently discovered non-photosynthetic cyanobacterial classes (*Sericytochromatia* and *Melainabacteria*).

Location: 237 locations across six continents encompassing multiple climates (arid, temperate, tropical, continental and polar) and vegetation types (forests, grasslands and shrublands).

Time period: Sampling was carried out between 2003 and 2015.

Major taxa studied: Photosynthetic and non-photosynthetic cyanobacterial taxa

Methods: We conducted a field survey and used co-occurrence network analysis and structural equation modelling to evaluate the distribution and environmental preferences of soil cyanobacteria across the globe. These ecological preferences were used to create a global atlas (predictive distribution maps) of soil cyanobacteria.

Results: Network analyses identified three major groups of cyanobacteria taxa, which resembled the three main cyanobacterial classes: the photosynthetic *Oxyphotobacteria*-dominated cluster, which were prevalent in arid and semiarid areas, and the non-photosynthetic

Sericytochromatia- and *Melainabacteria*-dominated clusters, which preferred hyperarid oligotrophic and acidic/ humid environments, respectively.

Main conclusions: This study provides novel insights into the environmental preferences of non-photosynthetic cyanobacteria in soils globally. Our findings highlight the contrasting environmental preferences among the three clusters of cyanobacteria and suggest that alterations in environmental conditions linked to climate change may result in important changes in the ecology and biogeography of these functionally important microorganisms.

Keywords: non-photosynthetic Cyanobacteria, Cyanobacteria, global distribution, microbial biogeography, microbial network, 16S amplicon sequencing

1 INTRODUCTION

Cyanobacteria are microorganisms responsible for some of the most important events in Earth's history, including the rise of oxygen levels via oxygenic photosynthesis (Dismukes *et al.*, 2001; Rasmussen *et al.*, 2008) and the formation of plastids through endosymbiosis (Mereschkowsky, 1905; Margulis, 1970). Despite being one of the most studied microbial groups (Castenholz *et al.*, 2001; Garcia-Pichel *et al.*, 2003; Garcia-Pichel, 2009; Whitton & Potts, 2012), there are still major gaps of knowledge associated with the diversity and global distribution of these organisms. Recent studies have revealed the existence of two new bacterial clades closely related to cyanobacteria, 4C0d-2 (*Melainabacteria*) and ML635J-21 (*Sericytochromatia*), recently proposed as new classes of phylum cyanobacteria (Soo *et al.*, 2014, 2017). These non-photosynthetic classes are included in the latest releases of the most commonly used rRNA databases, Silva and Greengenes (DeSantis *et al.*, 2006; Quast *et al.*, 2013). Unlike photosynthetic cyanobacteria (hereafter class *Oxyphotobacteria)*, these clades have no genes associated with photosynthesis, and have provided a new perspective on the phylum, broadening our understanding of the functional capabilities of cyanobacteria and their evolutionary origin.

The construction of metagenome-assembled genomes has enabled the assessment of the metabolic potential of these organisms, suggesting that *Melainabacteria* and *Sericytochromatia* are chemoheterotrophs with metabolisms mostly centered on fermentation (Di Rienzi *et al.*, 2013; Soo *et al.*, 2014, 2017; Soo, 2015). Additionally, no genes for phototrophy or carbon (C) fixation have been found in *Melainabacteria* and *Sericytochromatia* (Soo *et al.*, 2017), indicating that oxygenic photosynthesis could be a trait acquired later in *Oxyphotobacteria* by horizontal gene transfer (Raymond *et al.*, 2002). Such physiological and

genetic differences might result in contrasting ecological preferences for these novel cyanobacterial taxa, but empirical evidence for this is lacking.

Soil-borne *Oxyphotobacteria* are widely distributed on the Earth (Garcia-Pichel *et al.*, 2003; Whitton & Potts, 2012; Moreira *et al.*, 2013) but they are specially predominant in hot arid and polar regions with sparse plant cover. They are an important component of biocrusts, soil surface communities dominated by lichens, mosses, cyanobacteria and associated microorganisms (Weber *et al.* 2016) and play key ecological roles in these environments by regulating critical soil processes such as nitrogen (N) and C fixation, soil stabilization and infiltration/runoff (Mager & Thomas, 2011; Sciuto & Moro, 2015). Other terrestrial cyanobacterial communities grow on the surface or inside rocks and soil (endolithic and subsoils forms), and are well adapted to dry conditions and high or low irradiation regimes (Warren-Rhodes *et al.*, 2006; Domínguez & Asencio, 2011; Puente-Sánchez *et al.*, 2018). The capacity of *Oxyphotobacteria* to stay dormant during long periods of time is also a fundamental characteristic of these organisms, which allow them to survive in extreme environments characterized by high or low temperatures, desiccation regimes or high ultraviolet radiation (Garcia-Pichel, 2009; Quesada & Vincent, 2012; Whitton & Potts, 2012).

Local and regional studies show that soil *Oxyphotobacteria* are generally considered to prefer neutral to alkaline pH for optimum growth (Brock, 1973; Whitton & Sinclair, 1975; Nayak & Prasanna, 2007). However, the global biogeography of soil *Oxyphotobacteria* has not been fully resolved due to the concentration of cyanobacterial research in particular regions, e.g. studies in western United States or the Antarctic continent (Garcia-Pichel *et al.* 2001; Namsaraev *et al.* 2010)(Garcia-Pichel *et al.*, 2003; Moreira *et al.*, 2013; Büdel *et al.*, 2016; Williams *et al.*, 2016) and the focus given to key and abundant taxa, such as *Microcoleus vaginatus* or the genus *Chroococcidiopsis* (Bahl *et al.*, 2011; Dvořák *et al.*, 2012), or specific habitats such as cold ecosystems and deserts (Jungblut *et al.* 2010; Bahl *et al.* 2011). There are clear gaps of knowledge of their distribution in certain regions of the world, such as South America (Büdel *et al.*, 2016). Despite their wide dispersal ability due to small size, aeolian transport and tolerance to desiccation and irradiation (Billi *et al.*, 2000; Kellogg & Griffin, 2006), and their often cosmopolitan distribution (Garcia-Pichel *et al.*, 1996; Taton *et al.*, 2006; Flombaum *et al.*, 2013), current knowledge suggests a more complex biogeography of these microorganisms that is likely to be also influenced by their phylogeny and historical legacies (Garcia-Pichel *et al.*, 1996, 2003; Taton *et al.*, 2006; Nayak & Prasanna, 2007; Flombaum *et al.*, 2013).

The ecology and biogeography of the non-photosynthetic cyanobacteria classes (*Melainabacteria* and *Sericytochromatia)* in soils is poorly known. Available information on these organisms comes from genomes from aphotic environments such as animal guts or subsurface

groundwater and artificial systems such as water treatment facilities and laboratory bioreactors (Ley *et al.*, 2005; Warnecke *et al.*, 2007; Yagi *et al.*, 2010; Di Rienzi *et al.*, 2013; Soo *et al.*, 2014; Utami *et al.*, 2018) and the scarce environmental studies correspond only to aquatic ecosystems such as lakes and algal biofilms (Monchamp *et al.*, 2018, 2019).

To advance our understanding of the biogeography and ecological preferences of soil photosynthetic and non-photosynthetic cyanobacteria, we used data from a global soil survey covering a wide diversity of climate, soil and vegetation types (Delgado-Baquerizo *et al.*, 2018). We expected the distinct ecological attributes of photosynthetic and non-photosynthetic cyanobacteria to be associated with very different environmental preferences. For example, we know that some *Oxyphotobacteria* have developed highly competitive adaptations to thrive in arid soils with low soil organic C and plant productivity (Lund, 1967; Whitton & Sinclair, 1975; Maestre *et al.*, 2015). In these environments, we expect *Oxyphotobacteria* to dominate due to their capacity to build protective sheath pigments and to fix atmospheric C and N, which can be an important ecological advantage. However, *Oxyphotobacteria* are also expected to appear in a wide variety of environmental conditions, including low light, low oxygen or even anoxygenic environments due to their enormous functional diversity (Stal & Moezelaar, 1997; Adams & Duggan, 1999; Garcia-Pichel, 2009; Puente-Sánchez *et al.*, 2018). Conversely, non-photosynthetic cyanobacteria rely on soil organic C pools to grow, which could translate into contrasting preferences related to soil nutrient availability. We expect to find groups of taxa co-occurring and sharing similar environmental preferences (hereafter *ecological clusters*) related to photosynthetic capability, habitat preferences and historical legacies.

2 MATERIALS AND METHODS

2.1 Global survey: Sites, soil collection, soil and molecular analyses

We used 16S rRNA gene amplicon sequencing data from a global survey of 237 locations (Fig. S1) across six continents encompassing multiple climates (arid, temperate, tropical, continental and polar) and vegetation types (forests, grasslands and shrublands) (Delgado-Baquerizo *et al.*, 2018). A composite soil sample (0-7.5 cm depth) was collected under the dominant vegetation at each surveyed location. A fraction of each sample was immediately frozen at -20^{o}C for molecular analyses; the other fraction was air-dried for chemical analyses. Sample collection of soils took place between 2003 and 2015. We do not expect differences in the timing of sample collection to largely affect our results for two main reasons. First, at the global scale seasonal variability is expected to be largely overcome by cross-biome variability (e.g., see Carini *et al.*, 2020 on the importance of spatial vs. temporal scales when analyzing soil microbial communities). To put it simple, a dryland and a boreal forest are so different that usually harbor

distinct microbial communities regardless of their seasonal variability. Second, we are using amplicon sequencing DNA-based analyses (see below), which characterize not only the active portion of cyanobacterial communities but also the dormant one at the moment of sampling (Li *et al.*, 2017). The soils sampled comprise a wide variety of physico-chemical properties, pH ranged from 4.04 to 9.21, texture of the fine fraction (%clay+silt) ranged from 1.4 to 92.0%, soil total organic carbon (OC) from 0.15 to 34.77%, soil total nitrogen (TN) from 0.02 to 1.57, C:N ratio (CN) ranged from 2.12 to 67.52 and soil total phosphorus (TP) from 75.10 to 4111.04 mg P kg^{-1} soil. These analyses were done using standard laboratory methods described in Delgado-Baquerizo *et al.* (2018).

Climatic variables (maximum and minimum temperature [MAXT, MINT], precipitation seasonality [inter-annual coefficient of variation in precipitation, PSEA] and mean diurnal temperature range [MDR]) were obtained for each site from the WorldClim database (Hijmans *et al.*, 2005). Aridity Index (precipitation/potential evapotranspiration) was obtained from the Global Potential Evapotranspiration database (Zomer *et al.*, 2008), which uses interpolations from WorldClim. The annual ultraviolet index (UV Index), a measure of the risk of UV exposition ranging from 0 (minimal risk) to 16 (extreme risk), was obtained for each site using data from the Aura satellite (Newman & McKenzie, 2011). Net aboveground primary productivity [ANPP] was estimated with satellite imagery using the Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Terra satellites (Justice *et al.*, 1998). This index provides a global measure of the greenness of the Earth for a given period (Pettorelli *et al.*, 2005). Here, we used monthly averaged values for NDVI for the sampling period between 2003 and 2015 (10 km resolution).

Microbial DNA was extracted using the PowerSoil DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) following manufacturer's instructions. DNA extracts were sequenced targeting the bacterial V3-V4 region using 16S rRNA gene primers 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACHVGGGTATCTAATCC) and the Illumina Miseq platform of the Next Generation Genome Sequencing Facility at Western Sydney University (Australia). Bioinformatic analyses were performed with a combination of QIIME (Caporaso *et al.*, 2010), USEARCH (Edgar, 2010) and UPARSE (Edgar, 2013). After merging of the reads, the primers were trimmed and sequences of low quality (expected error rate > 1) were discarded. Phylotypes were defined with UCLUST (Edgar, 2010) at an identity level of 97% and taxonomy was assigned using Silva Incremental Alligner *Search and classify* with Silva database (complementing not identified phylotypes with Greengenes database) (DeSantis *et al.*, 2006; Quast *et al.*, 2013). Phylotypes represented by only a single read (singletons) were removed. The final dataset of phylotypes was filtered for phylum Cyanobacteria (excluding Chloroplast) and

the relative abundance each of cyanobacterial phylotype in relation to total bacteria (all 16S rRNA reads) was calculated.

2.2 Structure of the community: Network analyses

To explore the different patterns of cyanobacterial co-occurrence across our samples, we conducted a network analysis with the CoNet software (Faust & Raes, 2016). This tool detects significant non-random patterns of co-occurrence using multiple correlation and dissimilarity measures. Two correlation coefficients (Pearson and Spearman) and dissimilarity distances (Bray-Curtis and Kullback Leiber) were used to obtain a more reliable network (Faust & Raes, 2012). When links were detected by more than one correlation/dissimilarity measure, they were considered as a single link. Samples were standardized prior to network analyses with the "col_norm" function, which divides each column by its sum, converting abundances in column-wise proportions. We computed the network with the top 1000 links for each measure and tested the statistical significance of each link with 1000 permutations and the function "shuffle rows" as the resampling strategy. Multiple testing was corrected by using Benjamini-Hochberg's procedure (Benjamini & Hochberg, 1995), keeping links with an adjusted merged p-value below 0.05. The final network was visualized with the interactive platform gephi (Bastian *et al.*, 2009). We obtained the ecological clusters with the function "fastgreedy" from the igraph package (Csárdi & Nepusz, 2006) in R version 3.4.0 (Team, 2013), and tested the statistical significance of modularity using 10000 random networks. Network analysis allowed us to divide the community between ecological clusters, that we used for further analysis. The relative abundance of each ecological cluster per sample was calculated by averaging the standardized (z-score) relative abundance of the phylotypes present within each ecological cluster. Thus, we obtained a balanced contribution of each cyanobacterial phylotype to the relative abundance of its ecological cluster. Note that the use of z-score standardization transforms relative abundances, and therefore negative values can be obtained.

2.3 Factors determining cyanobacterial global distribution

Environmental effects: We conducted Structural Equation Modelling (SEM, Grace 2006) to evaluate the direct and indirect effects of spatial, climatic, vegetation and soil variables as predictors of the abundance of the main cyanobacterial ecological clusters (See Fig. S2 for our *a priori* model). This approach is useful for simultaneously testing the influence of multiple variables and the separation of direct and indirect effects of the predictors included in the model (Grace, 2006). These included spatial (Latitude, sine Longitude, cosine Longitude), climatic (MDR, MAXT, MINT, PSEA and Aridity [1-Aridity Index]) and vegetation (Grassland, Forest and

ANPP) variables, as well as soil properties (CN, soil OC, pH and percentage of clay and silt). Prior to modelling, we transformed them to improve normality: Aridity, OC, PSEA and CN were log-transformed and both ANPP and the percentages of clay and silt were square root transformed. We used the chi-square fit test, supplemented with root mean square error of approximation (RMSEA) to test the overall fit of the model. We analysed path coefficients of the model and their associated P values and the total effects of each variable. As some of the variables were not normally distributed despite transforming them, we used 5000 bootstraps to simultaneously test the significance of each path. SEM analyses were conducted using AMOS 24.0.0 (IBM SPSS, Chicago, IL, USA).

To obtain a prediction of the potential distribution of the main cyanobacterial ecological clusters, we used the regression model Cubist (Quinlan, 2014) as implemented in the R package Cubist (Kuhn *et al.*, 2016). This model uses a linear regression tree analysis that predicts the most important factors affecting the abundance of each ecological cluster based on environmental covariates. Covariates in our models included the same variables used in our SEMs. Global predictions of the distribution of major clusters were done on a 25 km resolution grid. Soil properties for this grid were obtained from SoilGrids (Hengl *et al.*, 2017). Major vegetation types (grasslands and forests) were obtained using Globcover2009 map from the European Space Agency (Bontemps *et al.*, 2013). Information on climate, UV index and net primary productivity were obtained from the WorldClim database and NASA satellites as described above.

We conducted multiple analyses to support the validity of our global prediction maps. First, we used kernel density estimations to compare the distribution of key soil and climate variables of our dataset with those from high resolution global maps: SoilGrids (Hengl *et al.*, 2017) and Worldclim (Hijmans *et al.*, 2005). Our dataset comprises a large percentage of their global variability (Fig. S3): 78.51% for OC, 94% for pH, 58.25% for Aridity, 45.98% for PSEA, 71.63% for MINT, 47.03% for MAXT and 96.43% for ANPP. These results indicate that our sampling covers a large proportion of the environmental variability found on Earth. Second, we found a strong correlation between the relative abundance of our cyanobacterial ecological clusters and key microbial environmental factors at the global scale (see results below), which suggests that environmental data can be used to predict their distribution. Finally, predictive maps were cross-validated with an independent dataset obtained from the Earth Microbiome Project (EMP, Thompson *et al.*, 2017), which contains data on soil cyanobacteria from 403 sites worldwide (see Fig. S1). For doing so, we estimated the relative abundance of the three main cyanobacterial clusters for the EMP dataset using the 97% similar EMP phylotypes. We first calculated relative abundance of each cyanobacterial phylotype in relation to total bacteria (all 16S rRNA reads of the EMP dataset). Then, the relative abundance of each ecological cluster per

sample was computed by averaging the standardized (z-score) relative abundance of the phylotypes of each ecological cluster, as explained above for our dataset. We then used our predictive maps to extract the predicted relative abundance of each cluster for the EMP locations. These predictive abundances were then compared with the independent results of relative abundance of each cluster calculated with the EMP dataset using Pearson correlations.

We also conducted a Permanova analysis with Bray Curtis distances to evaluate the effect of vegetation type on the abundance of each cyanobacterial cluster with the *adonis* function and 1000 permutations. To test for the differences in the relative abundance of each cluster across vegetation types we first tested the homogeneity of groups dispersions (variances) with *betadisper* function and from the result of that call we performed the post hoc analysis Tukey Honest Significant Differences with *TukeyHSD* function. All these analysis were done with vegan v2.4-2 (Oksanen, 2015) and R version 3.6.0 (Team, 2013).

Phylogenetic tree**:** The phylogenetic tree of cyanobacteria was constructed using the SILVA Alignment, Classification and Tree (ACT) Service (www.arb-silva.de/act). Multiple sequence alignment of the 343 rRNA gene sequences was performed using SINA v1.2.11 (Pruesse *et al.*, 2012). A phylogenetic tree was obtained with their built-in tree computation tool FastTree (Price *et al.*, 2009) using the General Time Reversible Model of nucleotide evolution (Nei & Kumar, 2000) and keeping the default parameters. The display and annotation of phylogenetic tree were made with iTol v5.5 (Letunic & Bork, 2019).

3 RESULTS

3.1 Global cyanobacterial co-occurrence patterns

Despite the common and widespread occurrence of soil cyanobacterial taxa on Earth, we did not find any of the 343 phylotypes present in all samples. The most ubiquitous cyanobacterial phylotype, *Microcoleus vaginatus*, was detected in 113 of the 237 sites surveyed. Moreover, the relative abundance of cyanobacterial phylotypes in our soils ranged from 0.01% to 4.35% of all bacterial 16S rRNA gene sequences (see Table S1). The cyanobacterial orders with the highest relative abundances included Oscillatoriales (*Oxyphotobacteria*), followed by Obscuribacterales (*Melainabacteria*) and Nostocales (*Oxyphotobacteria*) (Fig. 1). Non-photosynthetic phylotypes appeared almost in all samples (235/237 samples 99.2%). Photosynthetic cyanobacteria phylotypes appeared in the majority of them (185/237, 78.1%).

Our final network had 281 phylotypes and was arranged in 10 ecological clusters. Among these clusters, we identified three major groups of taxa co-occurring and comprising 65% of the

cyanobacterial phylotypes identified (Fig. 2a). The remaining seven clusters were minor, encompassing from 8% to 1% of phylotypes. The three main ecological clusters were dominated by either *Oxyphotobacteria* (82% of 76 phylotypes), *Sericytochromatia* (52% of 31 phylotypes) or *Melainabacteria* (83% of 76 phylotypes; see Table S1). We focused on these main ecological clusters for the downstream analyses. Our correlation network showed a contrasting node distribution for cyanobacterial phylotypes characterized by photosynthetic and non-photosynthetic capabilities (Fig. 2b). Overall, the three ecological clusters identified were strongly dominated by the three extant cyanobacterial classes (Fig. 2c, 2d).

3.2 Environmental preferences of photosynthetic and non-photosynthetic soil cyanobacteria

Vegetation type significantly affected the abundance of each of the main cyanobacterial clusters identified (Permanova R^2=0.28, 0.24 and 0.15 for *Melainabacteria*, *Sericytochromatia* and *Oxyphotobacteria*-dominated clusters, respectively, $p<0.05$ in all cases).

Our SEM model indicated that the cluster dominated by *Oxyphotobacteria* was positively and negatively related to aridity and net aboveground productivity, respectively (Figs. 3, 4 and S4a), which explains their high relative abundance in dry grasslands (Fig. 6). We also observed a positive association between the relative abundance of the *Oxyphotobacteria* dominated cluster and both soil pH and minimum temperature (Fig. 3, 4, and S4a). We predicted the distribution of this cluster in a wide range of arid and semiarid areas worldwide (e.g., southern Sahara, southern Africa, northern Australia, India, Arabian Peninsula, areas surrounding the Amazon Basin, southwestern US and northwestern Mexico; Fig. 5a).

The cluster dominated by *Sericytochromatia* had a strong preference for arid environments with low soil C content (Fig. 3, 4, 6 and S4b). Taxa within this ecological cluster were also positively associated with locations characterized by high inter-annual rainfall variability (Figs. 3, 4 and S4b). Our global atlas predicts that taxa within this ecological cluster can be found in hyper-arid areas such as the Saharan Desert, central Australia, the Atacama, Gobi and Taklamakan Deserts and the Arabian Peninsula, with almost no areas of intermediate relative abundance (Fig. 5b).

Unlike the other two ecological clusters identified, the *Melainabacteria*-dominated cluster showed a preference for humid and acidic soils, as indicated by the reduced relative abundance of this cluster with increases in aridity and pH (Figs. 3, 4 and. S4c). The vast majority of phylotypes found in our study corresponded to the order Obscuribacterales (1, 2d). This ecological cluster is found mainly in tropical and cold forests and grasslands (which are mostly temperate; see Fig. 6). Prediction maps show high relative abundance values of this cluster in

humid areas of the Amazon Basin, central Africa, west Asian coast and Pacific Islands (Fig 5c). Despite the methodological differences between our dataset and the EMP dataset (primer sets used here 341F/805R vs. 515F/806R for the EMP; read lengths here 400bp/sequence vs. <150bp for the EMP and the lack of standardization in the EMP soil sampling protocols and metadata collection) we obtained positive and significant correlations between both results: Melainabacteria dominated cluster Pearson's r=0.28 (P<0.001), Sericytochromatia dominated cluster Pearson's r=0.53 (P<0.001), Oxyphotobacteria dominated cluster Pearson's r=0.35 (P<0.001). These results support the validity of our maps as representative of the distribution of the main ecological clusters of cyanobacteria across the globe.

4 DISCUSSION

The discovery of non-photosynthetic cyanobacteria has expanded one of the currently most diverse bacterial phylum (Castenholz *et al.*, 2001; Garcia-Pichel, 2009; Whitton & Potts, 2012; Dvořák *et al.*, 2017). There is a large body of knowledge about photosynthetic cyanobacteria showing their importance in terrestrial ecosystems, as they are key components of cryptogamic covers, which are estimated to fix 3.9 Pg carbon per year (Elbert *et al.*, 2012). They increase soil fertility by fixing atmospheric N (Cleveland *et al.*, 1999), stabilize soils by producing extracellular polysaccharides (Mazor *et al.*, 1996; Mager & Thomas, 2011), protecting it from erosion and creating suitable habitats for the colonization of mosses and lichens (Zhang, 2005; Lan *et al.*, 2015). However we know relatively little about the distribution and environmental drivers of the newly described non-photosynthetic cyanobacteria in soils. Our work provides novel insights into the ecology and biogeography of these key organisms, and advances our understanding of on the potential vulnerabilities of photosynthetic and non-photosynthetic cyanobacteria to changing environmental conditions.

Photosynthetic taxa represented by the *Oxyphotobacteria*-dominated cluster prefer areas with sparse vegetation cover, and therefore greater accessibility to light, such as dry grasslands (Figs. 3,4, 6 and S4a). Accordingly, they are reported as key components of biocrust communities in low productivity ecosystems such as arid environments (Garcia-Pichel, 2009; Belnap *et al.*, 2016), where the ability to fix atmospheric C and N can be an important ecological advantage. As with the remaining bacterial communities (Fierer & Jackson, 2006) soil acidity is a key factor shaping the global distribution of *Oxyphotobacteria* (Fig. 4). Consistentwith previous studies (Baas-Becking *et al.*, 1960; Brock, 1973; Nayak & Prasanna, 2007) we found that photosynthetic cyanobacteria have a preference for neutral to alkaline soils (Figs. 3,4 and S4a), which are characteristic of drylands (Schlesinger & Bernhardt, 2013). Our analyses further indicate a wide distribution of this cluster in drylands worldwide (Fig. 5), as previously reported

for members of this taxa in continental-scale distribution studies (Bahl *et al.*, 2011; Garcia-Pichel *et al.*, 2013). Together with temperature, soil moisture plays a key role driving the physiology, small-scale distribution and behaviour of soil photosynthetic cyanobacteria (Garcia-Pichel & Pringault, 2001; Rajeev *et al.*, 2013). The high tolerance and photosynthetic performance of *Oxyphotobacteria* at high temperatures is one of the reasons why cyanobacterial-dominated biocrusts are so abundant in hyper-arid and arid environments (Grote *et al.*, 2010; Wang *et al.*, 2012). Thus, we observed a positive influence of high minimum temperatures and aridity on this cyanobacterial cluster (Figs. 3. and S4a). By moving from local/regional to the global scale, including samples from poorly-studied regions of South America (Garcia-Pichel *et al.*, 2003; Büdel *et al.*, 2016), and considering multiple terrestrial global biomes, our results provide novel predictions of the global distribution of *Oxyphotobacteria* in global soils.

Unlike *Oxyphotobacteria*, non-photosynthetic cyanobacteria require relatively large soil organic C pools for growth. We observed contrasting environmental preferences for each of the non-photosynthetic clusters across the oligotrophic-copiotrophic continuum, such as those reported for other soil heterotrophic organisms (e.g., methanotrophs in Nazaries *et al.* 2018). A key finding of our study is that the *Melainabacteria*-dominated cluster was especially abundant in mesic forests (tropical and cold forests, Fig. 6) and temperate grasslands, while the *Sericytochromatia*-dominated cluster is associated with locations with reduced plant cover and high temperatures (e.g., hyperarid deserts in Fig. 5, dry grasslands in Fig. 6). We found very little overlap between the predicted distributions of non-photosynthetic clusters of cyanobacteria (Figs. 5b, 5c) and a negative relationship between the relative abundances of these two non-photosynthetic clusters (Spearman correlation $r = -0.31$, $p<0.05$). Interestingly, a sizable percentage of members of *Melainabacteria* appears in the *Sericytochromatia* dominated-cluster (38%). We know that members of class *Melainabacteria* are capable of aerobic respiration because they contain respiratory components of the complex III-IV operon, which is adapted to low oxygen conditions, a C-family oxygen reductase and two cytochrome bc oxydases (Soo *et al.*, 2017). However, the *Melainabacteria*-dominated cluster is dominated by members of the order Obscuribacterales (Fig. 2d), for which there is little functional information available in the literature. Genomic analyses of the *Candidatus Obscuribacter phosphatis* suggest that this particular species is adapted to dynamic environments involving feast-famine nutrient cycles, and has the capacity for aerobic or anaerobic respiration and fermentation (Soo *et al.*, 2014). These features allow it to survive in both oxic and anoxic environments. To our knowledge there is no information available of the contribution of this cyanobacterium to the structure and function of forest ecosystems. However, our results suggest that molecular ecologists and taxonomists targeting taxa in *Melainabacteria*-dominated cluster should focus mainly on mesic

367 forests across the globe. We also expect non-photosynthetic cyanobacteria to play a significant
368 role in soil biogeochemical cycles in both high and low productive soils through C degradation
369 and/or H_2 production, as reported for *Melainabacteria* in an alluvial aquifer (Wrighton *et al.*,
370 2014). However, studies linking non-photosynthetic soil cyanobacteria to carbon degradation in
371 terrestrial environments are still lacking. Future studies are thus needed to identify the relative
372 contributions of non-photosynthetic cyanobacteria to organic matter decomposition and C
373 cycling in soils from contrasting biomes.
374 The topology of our phylogenetic tree (Fig. 2c) reflects the expected evolutionary
375 relationships from previous research with separation of three main clades (Soo *et al.*, 2017); the
376 basal deep branched *Sericytochromatia*, *Melainabacteria* and photosynthetic
377 *Oxyphotobacteria*. As the ecological clusters are related to these classes, their global distribution
378 is likely to be related to past evolutionary events within this ancient phylum (Bahl *et al.*, 2011;
379 Moreira *et al.*, 2013). The ecological diversification observed in the non-photosynthetic clades
380 is particularly noteworthy. We found a niche-differentiation between the basal cyanobacterial
381 clade, *Sericytochromatia,* which occupies extremely dry environments, and *Melainabacteria*,
382 which is mostly found in humid forests. Interestingly, the presence of phylotypes from
383 *Melainabacteria* in the *Sericytochromatia*-dominated cluster may point to the existence of
384 common ancestral traits between both classes and the later expansion of *Melainabacteria* into
385 new "humid" niches. Photosynthetic cyanobacteria (*Oxyphotobacteria*) are known for being
386 extraordinarily ecologically versatile, mostly living in environments with at least some exposure
387 to sunlight, and capable of inactivating their photosynthetic apparatus (Harel *et al.*, 2004) or
388 performing light-independent energy generation (Stal, 2012) when needed. There is still no
389 consensus about the date the acquisition of oxygenic photosynthesis by *Oxyphotobacteria*; this
390 could have happened either after divergence from other non-photosynthetic clades (Soo *et al.*,
391 2017) or before, sharing a photosynthetic common ancestor (Harel *et al.*, 2015). Regardless, the
392 acquisition of oxygenic photosynthesis was a revolutionary event that allowed cyanobacteria to
393 expand into diverse niches, and also the evolution of algae and terrestrial plants through
394 endosymbiosis (Mereschkowsky, 1905; Margulis, 1970).
395 Our findings represent a starting point towards the understanding of the ecological
396 preferences and global distributions of non-photosynthetic soil cyanobacteria. They highlight
397 the fact that major photosynthetic and non- photosynthetic groups of soil cyanobacteria have
398 contrasting ecological preferences across the globe. However, and given the difficulty of
399 predicting microorganisms at a global scale, conclusions should be viewed as preliminary. The
400 potential distribution maps presented here and the identification of the main environmental
401 drivers of soil cyanobacterial distribution also illustrate how different cyanobacterial lineages

might respond to ongoing climate and land use change. For example, the positive influence of aridity on the *Sericytochromatia*- and *Oxyphotobacteria*-dominated clusters suggests that the distribution of these taxa could expand under future climate change scenarios (Huang *et al.*, 2016). Consequently, our findings advance our understanding of the ecological distributions of these functionally important microbial communities and provide a basis for predicting possible future shifts of cyanobacterial terrestrial communities in a human-dominated, warmer and more arid world. To complement and expand our findings, future studies should further investigate the temporal dynamics of photosynthetic and non-photosynthetic cyanobacteria in terrestrial ecosystems, particularly along multiple temporal scales.

REFERENCES

Adams, D.G. & Duggan, P.S. (1999) Heterocyst and akinete differentiation in cyanobacteria. *New Phytologist*, **144**, 3–33.

Baas-Becking, L.G.M., Kaplan, I.R. & Moore, D. (1960) Limits of the natural environment in terms of pH and oxidation-reduction potentials. *The Journal of Geology*, **68**, 243–284.

Bahl, J., Lau, M.C.Y., Smith, G.J.D., Vijaykrishna, D., Cary, S.C., Lacap, D.C., Lee, C.K., Papke, R.T., Warren-Rhodes, K.A., Wong, F.K.Y., McKay, C.P. & Pointing, S.B. (2011) Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nature Communications*, **2**, 161–166.

Bastian, M., Heymann, S. & Jacomy, M. (2009) Gephi: an open source software for exploring and manipulating networks. *Proceedings of the 3rd International ICWSM Conference*, **8**, 361–362.

Belnap, J., Weber, B. & Büdel, B. (2016) *Biological Soil Crusts as an Organizing Principle in Drylands*. pp. 3–13.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300.

Billi, D., Friedmann, E.I., Hofer, K.G. & Caiola, M.G. (2000) Ionizing-radiation resistance in the desiccation-tolerant cyanobacterium *Chroococcidiopsis*. *Applied and Environmental Microbiology*, **66**, 1489–1492.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G., Zülkhe, M., Kalogirou, V. & Arino, O. (2013) *Consistent global land cover maps for climate modeling communities: Current*

achievements of the ESA's land cover CCI. *ESA Living Planet Symposium 9*,.

Brock, T.D. (1973) Lower pH limit for the existence of blue-green algae : Evolutionary and ecological implications. *Science*, **179**, 480–483.

Büdel, B., Dulić, T., Darienko, T., Rybalka, N. & Friedl, T. (2016) *Cyanobacteria and algae of Biological Soil Crusts*. *Biological Soil Crusts: An Organizing Principle in Drylands* (ed. by B. Weber), B. Büdel), and J. Belnap), pp. 55–80. Springer International Publishing, Cham.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335.

Carini, P., Delgado-Baquerizo, M., Hinckley, E.S., Brewer, T.E., Rue, G., Vanderburgh, C., Mcknight, D. & Fierer, N. (2020) Effects of Spatial Variability and Relic DNA Removal on the Detection of Temporal Dynamics in Soil Microbial. *Ecological and Evolutionary Science*, **11**, 1–16.

Castenholz, R.W., Wilmotte, A., Herdman, M., Rippka, R., Waterbury, J.B., Iteman, I. & Hoffmann, L. (2001) *Phylum BX. Cyanobacteria*. *Bergey's Manual of Systematic Bacteriology. Volume One : The Archaea and the Deeply Branching and Phototrophic Bacteria* (ed. by D.R. Boone), R.W. Castenholz), and G.M. Garrity), pp. 473–599. Springer New York, New York, NY.

Cleveland, C.C., Townsend, A.R., Schimel, D.S., Fisher, H., Hedin, L.O., Perakis, S., Latty, E.F., Fischer, C. Von, Elseroad, A. & Wasson, M.F. (1999) Global patterns of terrestrial biological nitrogen (Nz) fixation in natural ecosystems. **13**, 623–645.

Csárdi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal Complex Systems*, **1695**, 1–9.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K. & Fierer, N. (2018) A global atlas of the dominant bacteria found in soil. *Science*, **325**, 320–325.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database

and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.

Dismukes, G.C., Klimov, V. V., Baranov, S. V., Kozlov, Y.N., DasGupta, J. & Tyryshkin, A. (2001) The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences*, **98**, 2170–2175.

Domínguez, S.G. & Asencio, A.D. (2011) Distribution of chasmoendolithic cyanobacteria in gypsiferous soils from semi-arid environments (SE Spain) by chemical and physical parameters. *Nova Hedwigia*, **92**, 1–27.

Dvořák, P., Casamatta, D.A., Hašler, P., Jahodářová, E., Norwich, A.R. & PoPoulíčková, A. (2017) *Diversity of the Cyanobacteria*. *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects* (ed. by P.C. Hallenbeck), pp. 1–492.

Dvořák, P., Hašler, P. & Poulíčková, A. (2012) Phylogeography of the Microcoleus vaginatus (Cyanobacteria) from three continents - A spatial and temporal characterization. *PLoS ONE*, **7**.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996.

Elbert, W., Weber, B., Burrows, S., Steinkamp, J., Büdel, B., Andreae, M.O. & Pöschl, U. (2012) Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nature Geoscience*, **5**, 459–462.

Faust, K. & Raes, J. (2016) CoNet app: inference of biological association networks using Cytoscape. *F1000Research*, **5**, 1–14.

Faust, K. & Raes, J. (2012) Microbial interactions: From networks to models. *Nature Reviews Microbiology*, **10**, 538–550.

Fierer, N. & Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 626–631.

Flombaum, P., Gallegos, J.L., Gordillo, R. a, Rincón, J., Zabala, L.L., Jiao, N., Karl, D., Li, W., Lomas, M., Veneziano, D., Vera, C., Vrugt, J. a & Martiny, a C. (2013) Present and future

global distributions of the marine Cyanobacteria Prochlrococcus and Synechococcus. *Pnas*, **110**, 9824–9829.

Garcia-Pichel, F. (2009) *Cyanobacteria*. *Encyclopedia of Microbiology* (ed. by T.M. Schmidt), pp. 107–124. Academic Press.

Garcia-Pichel, F., Belnap, J., Neuer, S. & Schanz, F. (2003) Estimates of global cyanobacterial biomass and its distribution. *Algological Studies*, **109**, 213–227.

Garcia-Pichel, F., López-Cortés, A. & Nübel, U. (2001) Phylogenetic and morphological diversity of cyanobacteria in soil desert crusts from the Colorado Plateau. *Applied and Environmental Microbiology*, **67**, 1902–1910.

Garcia-Pichel, F., Loza, V., Marusenko, Y., Mateo, P. & Potrafka, R.M. (2013) Temperature drives the continental-scale distribution of key microbes in topsoil communities. *Science*, **340**, 1574–1577.

Garcia-Pichel, F. & Pringault, O. (2001) Cyanobacteria track water in desert soils. *Nature*, **413**, 380–381.

Garcia-Pichel, F., Prufert-Bebout, L. & Muyzer, G. (1996) Phenotypic and phylogenetic analyses show *Microcoleus chthonoplastes* to be a cosmopolitan cyanobacterium. *Applied and Environmental Microbiology*, **62**, 3284–3291.

Grace, J.B. (2006) *Structural equation modeling and natural systems*, Cambridge University Press.

Grote, E.E., Belnap, J., Housman, D. & Sparks, J.P. (2010) Carbon exchange in biological soil crust communities under differential temperatures and soil water contents : implications for global change. *Global Change Biology*, **16**, 2763–2774.

Harel, A., Karkar, S., Falkowski, P.G., Harel, A., Karkar, S. & Cheng, S. (2015) Deciphering primordial cyanobacterial genome functions from protein network analysis. *Current Biology*, **25**, 628–634.

Harel, Y., Ohad, I. & Kaplan, A. (2004) Activation of photosynthesis and resistance to photoinhibition in cyanobacteria within biological desert crust. *Plant Physiology*, **136**, 3070–3079.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara,

M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S. & Kempen, B. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, **12**, e0169748.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Huang, J., Yu, H., Guan, X., Wang, G. & Guo, R. (2016) Accelerated dryland expansion under climate change. *Nature Climate Change*, **6**, 166–171.

Justice, C.O., Vermote, E., Defries, R. & Roy, D.P. (1998) The Moderate Resolution Imaging Spectroradiometer (MODIS): Land Remote Sensing for Global Change Research. *IEEE transactions on geoscience and remote sensing*, **36**, 1228–1249.

Kellogg, C.A. & Griffin, D.W. (2006) Aerobiology and the global transport of desert dust. *Trends in Ecology and Evolution*, **21**, 638–644.

Kuhn, M., Weston, S., Keefer, C., Coulter, N. & Quinlan, R. (2016) Cubist: Rule-and Instance-Based Regression Modeling. R package version 0.0. 19.

Lan, S., Wu, L., Zhang, D. & Hu, C. (2015) Analysis of environmental factors determining development and succession in biological soil crusts. *Science of the Total Environment*, **538**, 492–499.

Letunic, I. & Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research*, **47**, W256–W259.

Ley, R.E., Backhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D. & Gordon, J.I. (2005) Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences*, **102**, 11070–11075.

Li, R., Tun, H.M., Jahan, M., Zhang, Z., Kumar, A., Fernando, D., Farenhorst, A. & Khafipour, E. (2017) Comparison of DNA-, PMA-, and RNA-based 16S rRNA Illumina sequencing for detection of live bacteria in water. *Scientific Reports*, **7**, 1–11.

Lund, J.W.G. (1967) *Soil algae*. *Soil biology* (ed. by A. Burges) and F. Raw), pp. 129–147. Elsevier.

Maestre, F.T., Delgado-Baquerizo, M., Jeffries, T.C., Eldridge, D.J., Ochoa, V., Gozalo, B., Quero, J.L., García-Gómez, M., Gallardo, A., Ulrich, W., Bowker, M.A., Arredondo, T., Barraza-

555 Zepeda, C., Bran, D., Florentino, A., Gaitán, J., Gutiérrez, J.R., Huber-Sannwald, E., Jankju,
556 M., Mau, R.L., Miriti, M., Naseri, K., Ospina, A., Stavi, I., Wang, D., Woods, N.N., Yuan, X.,
557 Zaady, E. & Singh, B.K. (2015) Increasing aridity reduces soil microbial diversity and
558 abundance in global drylands. *Proceedings of the National Academy of Sciences of the*
559 *United States of America*, **112**, 15684–15689.

560 Mager, D.M. & Thomas, A.D. (2011) Extracellular polysaccharides from cyanobacterial soil
561 crusts: A review of their role in dryland soil processes. *Journal of Arid Environments*, **75**,
562 91–97.

563 Margulis, L. (1970) *Origin of eukaryotic cells: Evidence and research implications for a theory of*
564 *the origin and evolution of microbial, plant and animal cells on the precambrian Earth*,
565 Yale University Press.

566 Mazor, G., Kidron, G.J., Vonshak, A. & Abeliovich, A. (1996) The role of cyanobacterial
567 exopolysaccharides desert microbial crusts. **21**, 121–130.

568 Mereschkowsky, C. (1905) Uber natur und ursprung der chromatophoren im pflanzenreiche.
569 *Biologisches Centralblatt*, **25**, 293–604.

570 Monchamp, M., Spaak, P., Domaizon, I., Dubois, N., Bouffard, D. & Pomati, F. (2018)
571 Homogenization of lake cyanobacterial communities over a century of climate change
572 and eutrophication. *Nature Ecology and Evolution*, **2**, 317–324.

573 Monchamp, M., Spaak, P. & Pomati, F. (2019) Long Term Diversity and Distribution of Non-
574 photosynthetic Cyanobacteria in Peri-Alpine Lakes. **9**, 1–11.

575 Moreira, C., Vasconcelos, V. & Antunes, A. (2013) Phylogeny and biogeography of
576 cyanobacteria and their produced toxins. *Marine Drugs*, **11**, 4350–4369.

577 Namsaraev, Z., Mano, M.J., Fernandez, R. & Wilmotte, A. (2010) Biogeography of terrestrial
578 cyanobacteria from Antarctic ice-free areas. *Annals of Glaciology*, **51**, 171–177.

579 Nayak, S. & Prasanna, R. (2007) Soil pH and its role in cyanobacterial abundance and diversity
580 in rice field soils. *Applied Ecology and Environmental Research*, **5**, 103–113.

581 Nazaries, L., Karunaratne, S.B., Delgado-Baquerizo, M., Campbell, C.D. & Singh, B.K. (2018)
582 Environmental drivers of the geographical distribution of methanotrophs: Insights from a
583 national survey. *Soil Biology and Biochemistry*, **127**, 264–279.

584 Nei, M. & Kumar, S. (2000) *Molecular evolution and phylogenetics*, Oxford University Press.

585 Newman, P.A. & McKenzie, R. (2011) UV impacts avoided by the Montreal Protocol.
586 *Photochemical & Photobiological Sciences*, **10**, 1152–1160.

587 Oksanen, J. (2015) Vegan: an introduction to ordination. *URL http://cran. r-project.*
588 *org/web/packages/vegan/vignettes/introvegan. pdf*, **8**, 19.

589 Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.M., Tucker, C.J. & Stenseth, N.C. (2005) Using
590 the satellite-derived NDVI to assess ecological responses to environmental change.
591 *Trends in Ecology and Evolution*, **20**, 503–510.

592 Price, M.N., Dehal, P.S. & Arkin, A.P. (2009) Fasttree: Computing large minimum evolution
593 trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, **26**,
594 1641–1650.

595 Pruesse, E., Peplies, J. & Glöckner, F.O. (2012) SINA: Accurate high-throughput multiple
596 sequence alignment of ribosomal RNA genes. *Bioinformatics*, **28**, 1823–1829.

597 Puente-Sánchez, F., Arce-Rodríguez, A., Oggerin, M., García-Villadangos, M., Moreno-Paz, M.,
598 Blanco, Y., Rodríguez, N., Bird, L., Lincoln, S.A., Tornos, F., Prieto-Ballesteros, O., Freeman,
599 K.H., Pieper, D.H., Timmis, K.N. & Amils, R. (2018) Viable cyanobacteria in the deep
600 continental subsurface. *Proceedings of the National Academy of Sciences*, **115**, 10702–
601 10707.

602 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F.O.
603 (2013) The SILVA ribosomal RNA gene database project: Improved data processing and
604 web-based tools. *Nucleic Acids Research*, **41**, 590–596.

605 Quesada, A. & Vincent, W.F. (2012) *Cyanobacteria in the cryosphere: snow, ice and extreme*
606 *cold*. *Ecology of cyanobacteria II* (ed. by B.A. Whitton), pp. 387–399. Springer.

607 Quinlan, J.R. (2014) *C4. 5: programs for machine learning*, Elsevier.

608 Rajeev, L., Nunes, U., Klitgord, N., Luning, E.G., Fortney, J., Axen, S.D., Shih, P.M., Bouskill, N.J.,
609 Bowen, B.P., Kerfeld, C.A., Garcia-pichel, F., Brodie, E.L., Northen, T.R. & Mukhopadhyay,
610 A. (2013) Dynamic cyanobacterial response to hydration and dehydration in a desert
611 biological soil crust. **7**, 2178–2191.

612 Rasmussen, B., Fletcher, I.R., Brocks, J.J. & Kilburn, M.R. (2008) Reassessing the first
613 appearance of eukaryotes and cyanobacteria. *Nature*, **455**, 1101–1104.

614 Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y. & Blankenship, R.E. (2002) Whole-

genome analysis of photosynthetic prokaryotes. *Science*, **298**, 1616–1620.

Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F. & Ley, R.E. (2013) The human gut and subsurface harbor non-photosynthetic Cyanobacteria. *Elife*, **2:e01102**, 1–25.

Schlesinger, W.H. & Bernhardt, E.S. (2013) *Biogeochemistry: an analysis of global change*, Academic press.

Sciuto, K. & Moro, I. (2015) Cyanobacteria: the bright and dark sides of a charming group. *Biodiversity and Conservation*, **24**, 711–738.

Soo, R.M. (2015) In search of non-photosynthetic Cyanobacteria.

Soo, R.M., Hemp, J., Parks, D.H., Fischer, W.W. & Hugenholtz, P. (2017) On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, **355**, 1436–1440.

Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., Steen, J.A., Parks, D.H., Tyson, G.W. & Hugenholtz, P. (2014) An expanded genomic representation of the phylum Cyanobacteria. *Genome Biology and Evolution*, **6**, 1031–1045.

Stal, L.J. (2012) *Cyanobacterial mats and stromatolites*. *Ecology of cyanobacteria II* (ed. by B.A. Whitton), pp. 65–125. Springer.

Stal, L.J. & Moezelaar, R. (1997) Fermentation in cyanobacteria. *FEMS Microbiology Reviews*, **21**, 179–211.

Taton, A., Grubisic, S., Balthasart, P., Hodgson, D.A., Laybourn-Parry, J. & Wilmotte, A. (2006) Biogeographical distribution and ecological ranges of benthic cyanobacteria in East Antarctic lakes. *FEMS Microbiology Ecology*, **57**, 272–289.

Team, R.C. (2013) R: A language and environment for statistical computing.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G., Navas-Molina, J.A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J.T., Mirarab, S., Xu, Z.Z., Jiang, L., Haroon, M.F., Kanbar, J., Zhu, Q., Song, S.J., Kosciolek, T., Bokulich, N.A., Lefler, J., Brislawn, C.J., Humphrey, G., Owens, S.M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J.A., Clauset, A., Stevens, R.L., Shade, A., Pollard, K.S., Goodwin, K.D., Jansson, J.K., Gilbert, J.A., Knight, R., Agosto Rivera, J.L., Al-Moosawi, L., Alverdy, J., Amato, K.R., Andras, J.,

Angenent, L.T., Antonopoulos, D.A., Apprill, A., Armitage, D., Ballantine, K., Bárta, J., Baum, J.K., Berry, A., Bhatnagar, A., Bhatnagar, M., Biddle, J.F., Bittner, L., Boldgiv, B., Bottos, E., Boyer, D.M., Braun, J., Brazelton, W., Brearley, F.Q., Campbell, A.H., Caporaso, J.G., Cardona, C., Carroll, J.L., Cary, S.C., Casper, B.B., Charles, T.C., Chu, H., Claar, D.C., Clark, R.G., Clayton, J.B., Clemente, J.C., Cochran, A., Coleman, M.L., Collins, G., Colwell, R.R., Contreras, M., Crary, B.B., Creer, S., Cristol, D.A., Crump, B.C., Cui, D., Daly, S.E., Davalos, L., Dawson, R.D., Defazio, J., Delsuc, F., Dionisi, H.M., Dominguez-Bello, M.G., Dowell, R., Dubinsky, E.A., Dunn, P.O., Ercolini, D., Espinoza, R.E., Ezenwa, V., Fenner, N., Findlay, H.S., Fleming, I.D., Fogliano, V., Forsman, A., Freeman, C., Friedman, E.S., Galindo, G., Garcia, L., Garcia-Amado, M.A., Garshelis, D., Gasser, R.B., Gerdts, G., Gibson, M.K., Gifford, I., Gill, R.T., Giray, T., Gittel, A., Golyshin, P., Gong, D., Grossart, H.P., Guyton, K., Haig, S.J., Hale, V., Hall, R.S., Hallam, S.J., Handley, K.M., Hasan, N.A., Haydon, S.R., Hickman, J.E., Hidalgo, G., Hofmockel, K.S., Hooker, J., Hulth, S., Hultman, J., Hyde, E., Ibáñez-Álamo, J.D., Jastrow, J.D., Jex, A.R., Johnson, L.S., Johnston, E.R., Joseph, S., Jurburg, S.D., Jurelevicius, D., Karlsson, A., Karlsson, R., Kauppinen, S., Kellogg, C.T.E., Kennedy, S.J., Kerkhof, L.J., King, G.M., Kling, G.W., Koehler, A. V., Krezalek, M., Kueneman, J., Lamendella, R., Landon, E.M., Lanede Graaf, K., LaRoche, J., Larsen, P., Laverock, B., Lax, S., Lentino, M., Levin, I.I., Liancourt, P., Liang, W., Linz, A.M., Lipson, D.A., Liu, Y., Lladser, M.E., Lozada, M., Spirito, C.M., MacCormack, W.P., MacRae-Crerar, A., Magris, M., Martín-Platero, A.M., Martín-Vivaldi, M., Martínez, L.M., Martínez-Bueno, M., Marzinelli, E.M., Mason, O.U., Mayer, G.D., McDevitt-Irwin, J.M., McDonald, J.E., McGuire, K.L., McMahon, K.D., McMinds, R., Medina, M., Mendelson, J.R., Metcalf, J.L., Meyer, F., Michelangeli, F., Miller, K., Mills, D.A., Minich, J., Mocali, S., Moitinho-Silva, L., Moore, A., Morgan-Kiss, R.M., Munroe, P., Myrold, D., Neufeld, J.D., Ni, Y., Nicol, G.W., Nielsen, S., Nissimov, J.I., Niu, K., Nolan, M.J., Noyce, K., O'Brien, S.L., Okamoto, N., Orlando, L., Castellano, Y.O., Osuolale, O., Oswald, W., Parnell, J., Peralta-Sánchez, J.M., Petraitis, P., Pfister, C., Pilon-Smits, E., Piombino, P., Pointing, S.B., Pollock, F.J., Potter, C., Prithiviraj, B., Quince, C., Rani, A., Ranjan, R., Rao, S., Rees, A.P., Richardson, M., Riebesell, U., Robinson, C., Rockne, K.J., Rodriguezl, S.M., Rohwer, F., Roundstone, W., Safran, R.J., Sangwan, N., Sanz, V., Schrenk, M., Schrenzel, M.D., Scott, N.M., Seger, R.L., Seguinorlando, A., Seldin, L., Seyler, L.M., Shakhsheer, B., Sheets, G.M., Shen, C., Shi, Y., Shin, H., Shogan, B.D., Shutler, D., Siegel, J., Simmons, S., Sjöling, S., Smith, D.P., Soler, J.J., Sperling, M., Steinberg, P.D., Stephens, B., Stevens, M.A., Taghavi, S., Tai, V., Tait, K., Tan, C.L., Taş, N., Taylor, D.L., Thomas, T., Timling, I., Turner, B.L., Urich, T., Ursell, L.K., Van Der Lelie, D., Van Treuren, W., Van Zwieten, L., Vargas-Robles, D., Thurber, R.V.,

680 Vitaglione, P., Walker, D.A., Walters, W.A., Wang, S., Wang, T., Weaver, T., Webster, N.S.,
681 Wehrle, B., Weisenhorn, P., Weiss, S., Werner, J.J., West, K., Whitehead, A., Whitehead,
682 S.R., Whittingham, L.A., Willerslev, E., Williams, A.E., Wood, S.A., Woodhams, D.C., Yang,
683 Y., Zaneveld, J., Zarraonaindia, I., Zhang, Q. & Zhao, H. (2017) A communal catalogue
684 reveals Earth's multiscale microbial diversity. *Nature*, **551**, 457–463.

685 Utami, Y.D., Kuwahara, H., Murakami, T., Morikawa, T., Sugaya, K., Kihara, K., Yuki, M., Lo, N.,
686 Deevong, P., Hasin, S., Boonriam, W., Inoue, T., Yamada, A., Ohkuma, M. & Hongoh, Y.
687 (2018) Phylogenetic diversity and single-cell genome analysis of "Melainabacteria", a
688 non-photosynthetic cyanobacterial group , in the termite gut. *Microbes and*
689 *Environments*, **33**, 50–57.

690 Wang, W., Wang, Y., Shu, X. & Zhang, Q. (2012) Physiological responses of soil crust-forming
691 cyanobacteria to diurnal temperature variation. *Journal of Basic Microbiology*, **52**, 1–9.

692 Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T.,
693 Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S.G., Podar,
694 M., Martin, H.G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov,
695 A., Barry, K., Mikhailova, N., Kyrpides, N.C., Matson, E.G., Ottesen, E.A., Zhang, X.,
696 Hernández, M., Murillo, C., Acosta, L.G., Rigoutsos, I., Tamayo, G., Green, B.D., Chang, C.,
697 Rubin, E.M., Mathur, E.J., Robertson, D.E., Hugenholtz, P. & Leadbetter, J.R. (2007)
698 Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher
699 termite. *Nature*, **450**, 560.

700 Warren-Rhodes, K.A., Rhodes, K.L., Pointing, S.B., Ewing, S.A., Lacap, D.C., Gómez-Silva, B.,
701 Amundson, R., Friedmann, E.I. & McKay, C.P. (2006) Hypolithic cyanobacteria, dry limit of
702 photosynthesis, and microbial ecology in the hyperarid Atacama Desert. *Microbial*
703 *Ecology*, **52**, 389–398.

704 Whitton, B. & Sinclair, C. (1975) Ecology of blue-green algae. *Science Reviews 2000 Ltd.*, **62**,
705 429–446.

706 Whitton, B.A. & Potts, M. (2012) *Introduction to the cyanobacteria*. *Ecology of Cyanobacteria II*
707 (ed. by B.A. Whitton), pp. 1–13. Springer.

708 Williams, L., Loewen-Schneider, K., Maier, S. & Büdel, B. (2016) Cyanobacterial diversity of
709 western European biological soil crusts along a latitudinal gradient. *FEMS Microbiology*
710 *Ecology*, **92**, fiw157.

711 Wrighton, K.C., Castelle, C.J., Wilkins, M.J., Hug, L.A., Sharon, I., Thomas, B.C., Handley, K.M.,
712 Mullin, S.W., Nicora, C.D., Singh, A., Lipton, M.S., Long, P.E., Williams, K.H. & Banfield, J.F.
713 (2014) Metabolic interdependencies between phylogenetically novel fermenters and
714 respiratory organisms in an unconfined aquifer. 1452–1463.

715 Yagi, J.M., Neuhauser, E.F., Ripp, J.A., Mauro, D.M. & Madsen, E.L. (2010) Subsurface
716 ecosystem resilience: Long-term attenuation of subsurface contaminants supports a
717 dynamic microbial community. *ISME Journal*, **4**, 131–143.

718 Zhang, Y. (2005) The microstructure and formation of biological soil crusts in their early
719 developmental stage. **50**, 117–121.

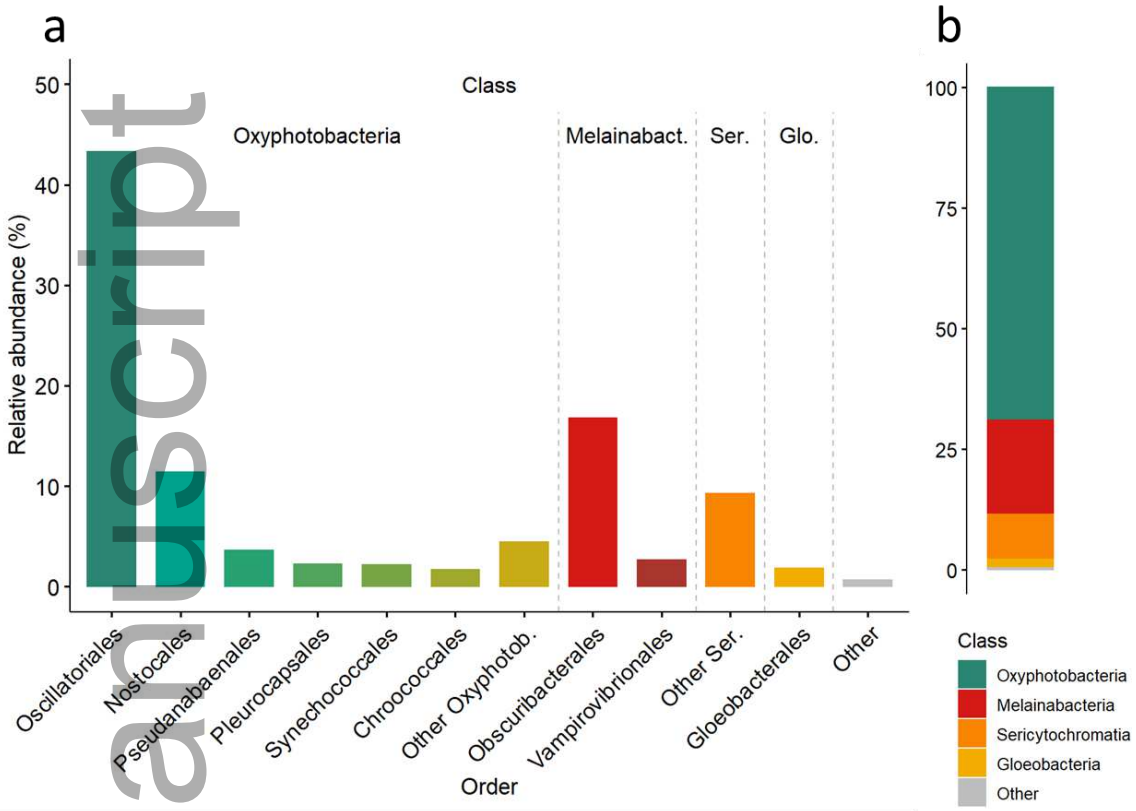
720 Zomer, R.J., Trabucco, A. & Bossio, D.A. (2008) Climate change mitigation: A spatial analysis of
721 global land suitability for clean development mechanism afforestation and reforestation.
722 *Agriculture, Ecosystems & Environment*, **126**, 67–80.

723

724 DATA ACCESSIBILITY STATEMENT

725 Raw data related with this manuscript are available in Figshare,
726 https://figshare.com/s/82a2d3f5d38ace925492

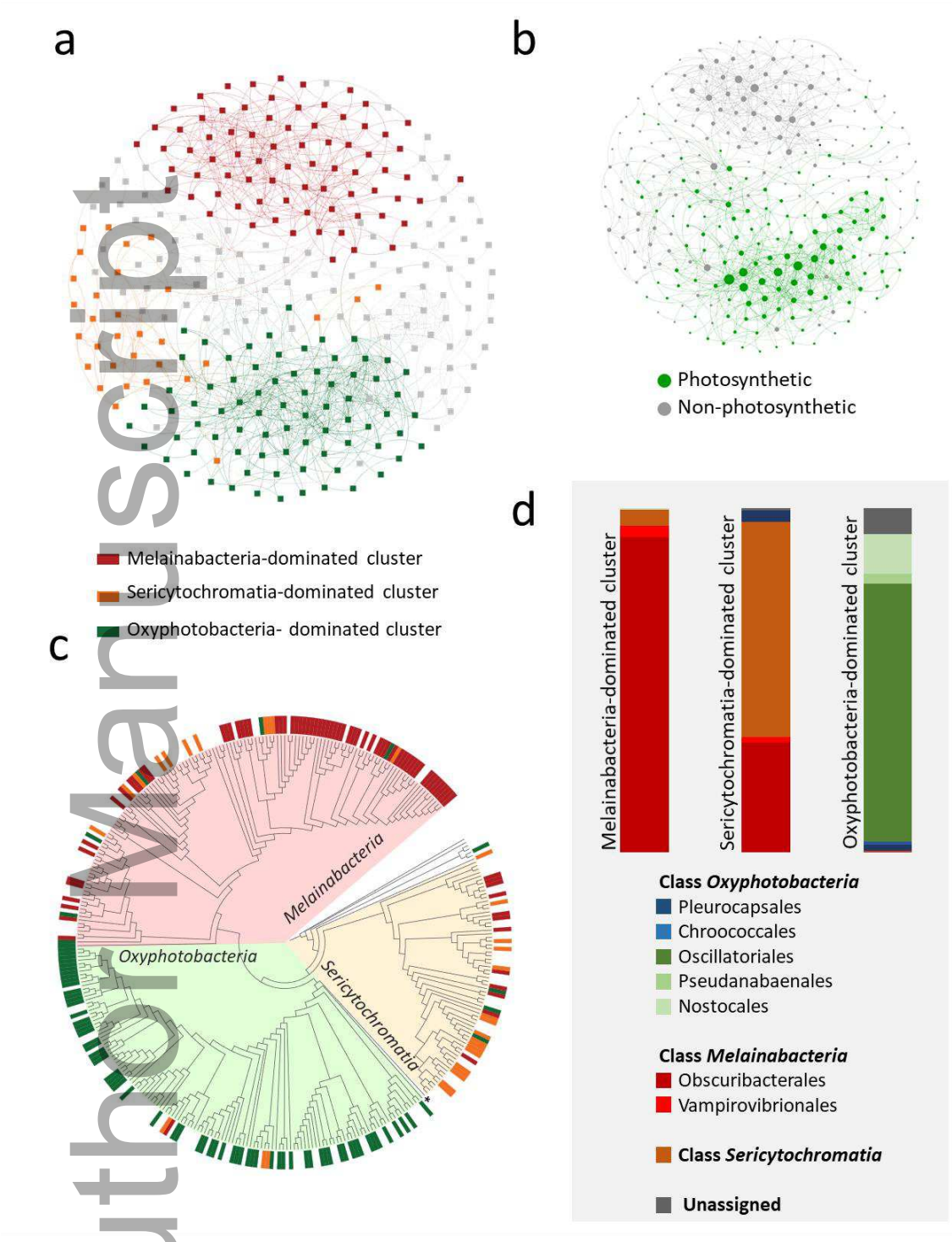
727 FIGURES



728

729 **Fig. 1.** Taxonomic information on the relative abundance of cyanobacterial orders (a) and classes
730 (b) across all sites. Ser.= Sericytochromatia (no orders described yet) and Glo. = Gloeobacteria.

731



732

733 **Fig. 2** Global network of co-occurrences within soil cyanobacteria, colored by either main
734 ecological clusters (a) or the photosynthetic capability of taxa (b). The size of the nodes is related
735 to the number of links they contain. The network had 282 nodes (cyanobacterial phylotypes)
736 and 986 significant links (potential ecological interactions between phylotypes) (c) Phylogenetic
737 tree obtained with the main ecological clusters located at the end of the branch. Background

colored by cyanobacterial class, * for Gloeobacteria class. (d) Taxonomic composition in relation to total 16S reads.

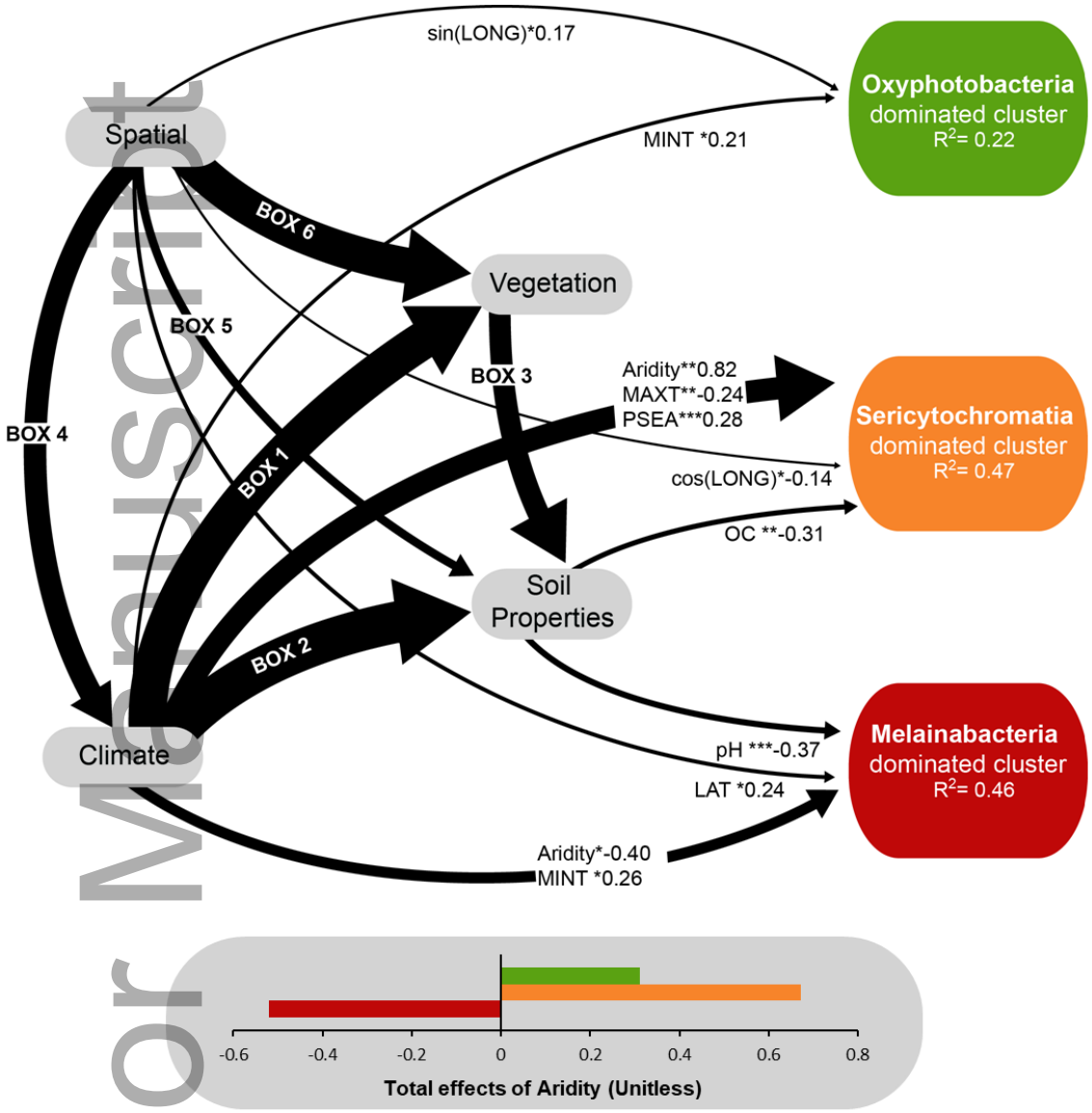
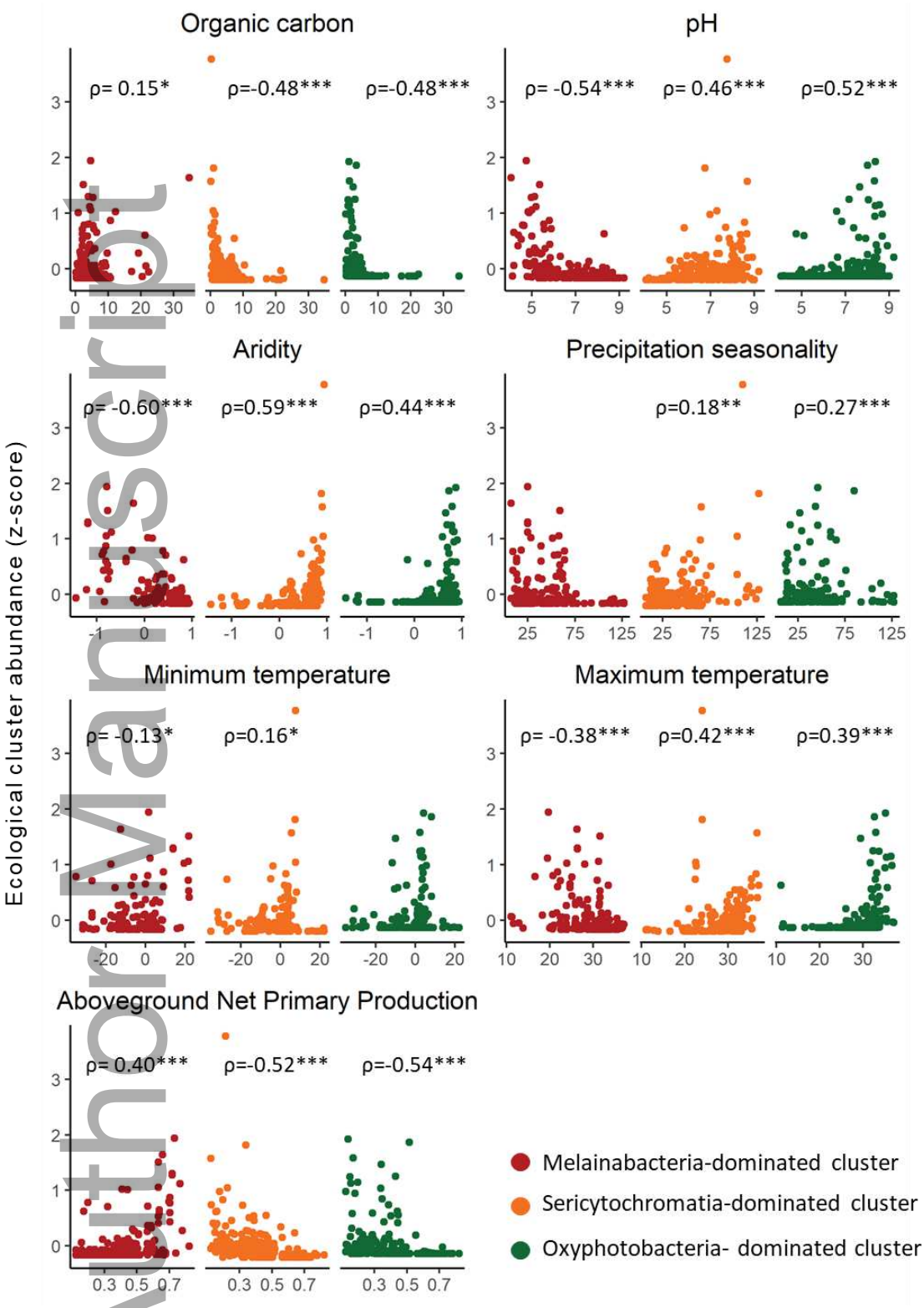


Fig. 3 Structural equation modelling (SEM) showing the direct effects of spatial (Latitude [LAT], Sine Longitude [sin(LONG)] and Cosine Longitude [cos(LONG)]) , climatic (maximum temperature [MAXT], minimum temperature [MINT], precipitation seasonality [PSEA] and aridity, calculated as 1-aridity index) and soil (soil organic carbon [OC] and pH) variables on the abundance of each ecological cluster. Numbers in arrows indicate standardized path coefficients, and their width is proportional to the strength of path coefficients. The proportion of variance explained (R^2) appears below every response variable in the model. Significance levels are as follows *P<0.05, **P<0.01, and ***P<0.001. Model X^2 =2.567, P= 0.463 df= 3, Bootstrap p= 0.254. Information on boxes 1-6 is shown in Fig. S2.



750

751 **Fig. 4** Relationships between main environmental predictors and the relative abundance (z-
752 score) of each one of the cyanobacterial clusters. Significant (P<0.05) spearman correlation
753 coefficients are shown on the upper part of each panel.

754

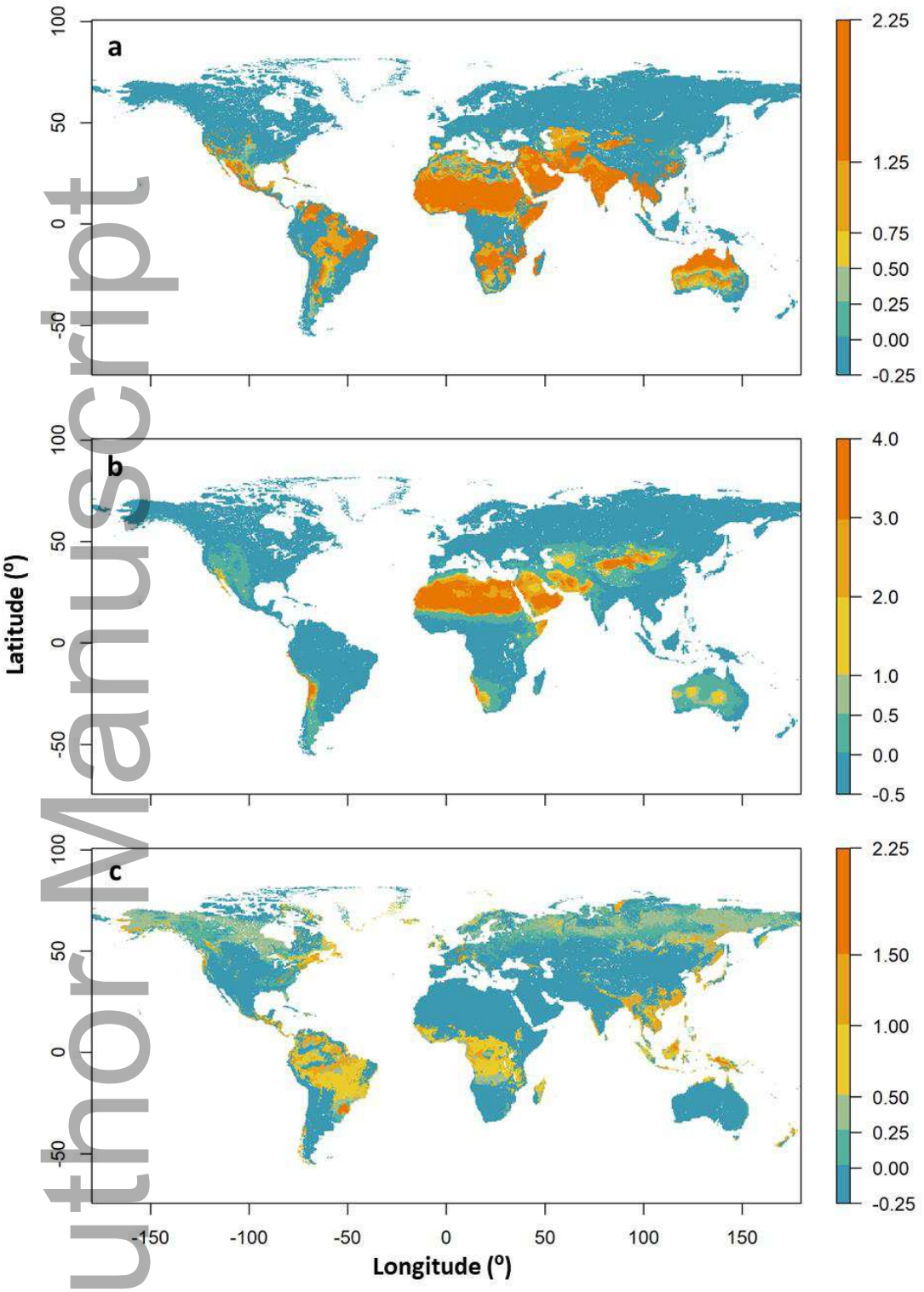
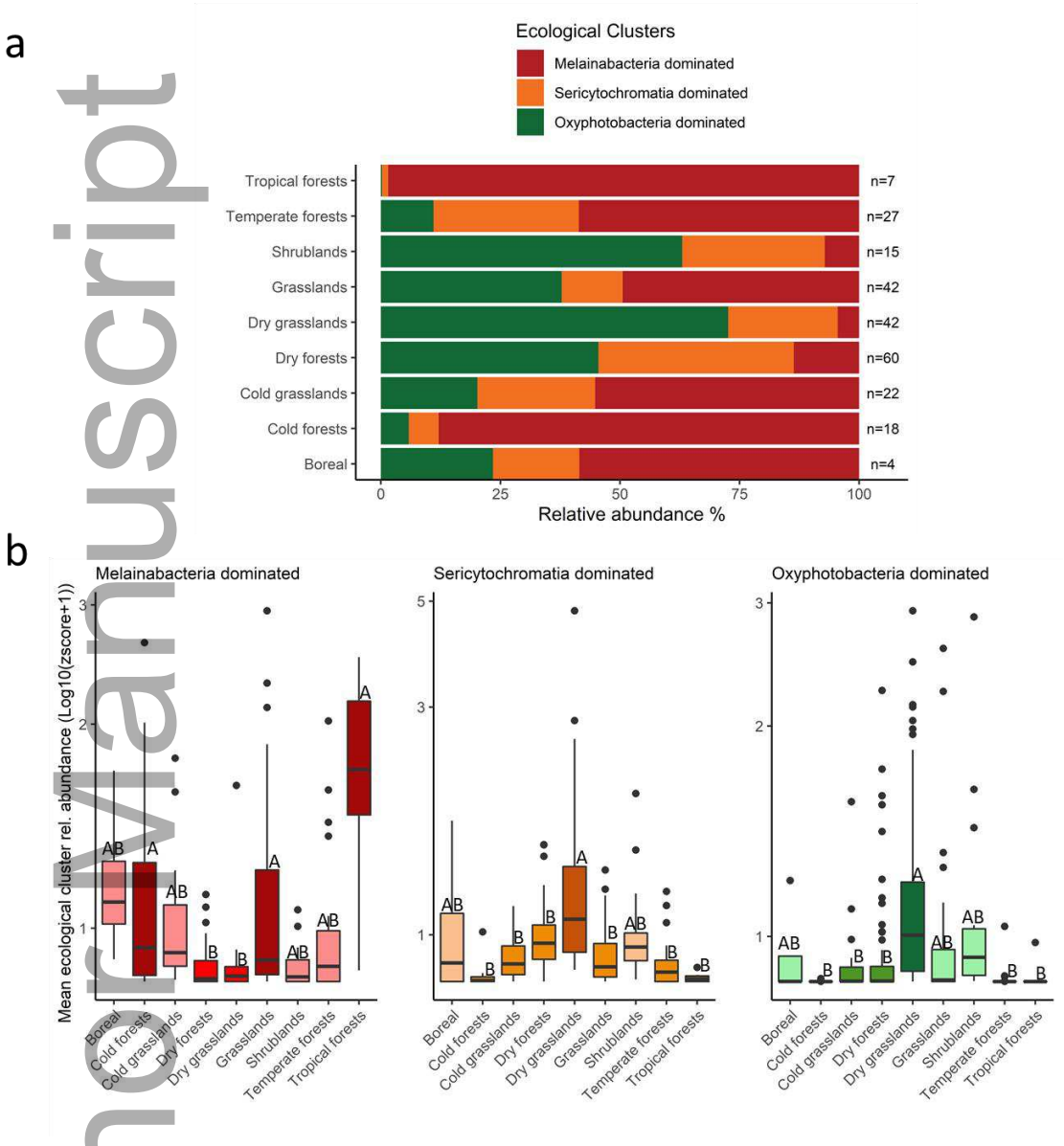


Fig. 5 Predicted global distribution of the relative abundance of the main ecological clusters of soil cyanobacteria. Percentage of variation explained by the models as follows: (a) *Oxyphotobacteria*-dominated cluster R^2 = 0.28; P < 0.001, (b) *Sericytochromatia*-dominated cluster R^2 =0.66; P < 0.001, (c) *Melainabacteria*-dominated cluster R^2 = 0.35; P < 0.001. The scale bar represents the standardized abundance (z-score) of each ecological cluster. An independent

761 cross-validation for these maps using data from the Earth Microbiome Project (Thompson *et al.*,
762 2017) is described in the Methods section.



763

764 **Fig. 6** Relative abundance of cyanobacterial clusters across major vegetation types. A) Stacked
765 bars showing the percentage of phylotypes of each ecological cluster per vegetation type.
766 n=Number of sites per each vegetation type B) Tukey HSD results testing the differences (letters
767 and colour hues) in the relative abundances of each ecological cluster across vegetation types.