Cross-biome drivers of soil bacterial alpha diversity on a worldwide scale.

Manuel Delgado-Baquerizo[1,2*], David J Eldridge[3].

1. Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309.

2. Departamento de Biología y Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Calle Tulipán Sin Número, Móstoles 28933, Spain.

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

***Authors for correspondence:**

Manuel Delgado-Baquerizo. Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309. E-mail: M.DelgadoBaquerizo@gmail.com

Statement of authorship: M.D-B. designed this study. M.D-B. and D.J.E. were involved in the collection of the soils used in this study. M.D-B. conducted statistical modeling. The manuscript was written by M.D-B, and edited by D.J.E.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 702057.

Abstract

We lack a defined suite of attributes that allow us to universally predict the distribution of bacterial diversity across and within globally distributed biomes. Using data from a global survey, including 237 locations and multiple environmental predictors, we found that only ultraviolet light, forest environments, soil carbon and pH can be considered as significant and globally consistent predictors of soil bacterial diversity, valid within and across biomes (arid, temperate and continental). Bacterial diversity always peaked in grasslands, with moderate to low carbon and ultraviolet light levels, and high soil pH. Using these environmental data, we generated the first global predictive map of the distribution of soil bacterial diversity. Our work helps to identify a unique set of environmental attributes for universally predicting the distribution of soil bacterial diversity. This knowledge is key to help predict changes in ecosystem functioning and the provision of essential services under changing environments.

Keywords. α-Diversity; Terrestrial ecosystems; Arid; Continental; Temperate; cross-biome.

Introduction

Biodiversity is critically important for the maintenance of ecosystem functions and services that are essential for human well-being (Hooper *et al.* 2005; Cardinale *et al.* 2012). Human activity over the past 500 years (the Anthropocene) has resulted in substantial reductions in biodiversity (Dirzo *et al.* 2014), which also affect the diversity of soil microorganisms (Maestre *et al.* 2015). Bacterial alpha diversity (the number of bacterial phylotypes) plays a critical role in driving multiple soil processes including nutrient cycling, litter decomposition, toxin degradation, gases emissions and plant productivity (Bodelier 2011; Fierer *et al.* 2012; Philippot *et al.* 2013; Jing *et al.* 2015; Delgado-Baquerizo *et al.* 2016a; 2017). We have a good understanding of the major environmental predictors of soil bacterial alpha diversity (hereafter, bacterial diversity) in terrestrial ecosystems from local to global scales (see Fierer 2017 for a recent review). These environmental factors include (1) soil properties such as pH (Lauber *et al.* 2009), carbon and nutrient content (Waldrop *et al.* 2006; Delgado-Baquerizo *et al.* 2017) (2) climatic factors such as rainfall amount (Maestre *et al.* 2015), air temperature (Zhou *et al.* 2016) and climatic variability (Delgado-Baquerizo *et al.* 2016b), and to a lesser extent, (3) plant attributes (Crowther *et al.* 2014; Prober *et al.* 2015). However, many others factors such as plant primary productivity and ultraviolet radiation have been largely neglected as global predictors of bacterial diversity; being all these key factors shaping plant biodiversity (Mackerness 2010; Isbell *et al.* 2011; Reich *et al.* 2012). Despite this knowledge, we still lack a defined suite of attributes that will allow us to universally predict the distribution of bacterial diversity across and within globally distributed biomes. We know that specific environmental factors such as soil pH are strong and consistent drivers of bacterial diversity. Soil pH has been shown to affect bacterial diversity across multiple terrestrial environments (Lauber *et al.* 2009; Rousk *et al.* 2010; Ramirez *et al.* 2014; Delgado-Baquerizo *et al.* 2016a). The study of other strong and consistent global predictors of bacterial alpha diversity across different biomes has, however, remained largely neglected, despite the fact that bacterial alpha diversity is intimately involved in many global environmental processes.

A clear and unambiguous suite of global predictors of bacterial diversity (valid across and within biomes), other than soil pH, remains elusive for several reasons. First, most previous studies have focused on specific regions on Earth. However, we know that the identity and importance of environmental predictors of bacterial alpha diversity often shift across biomes. For example, Maestre *et al.* (2015) found that aridity is the major driver of bacterial alpha diversity in global

drylands. Similarly, Zhou *et al.* (2016) suggested that air temperature mediates continental-scale diversity of microbes in forest soils from North America. In temperate ecosystems from Scotland, Delgado-Baquerizo *et al.* (2017) demonstrated that soil nutrient content was a strong predictor of bacterial diversity in soils. Thus the widespread applicability of these environmental factors as within and cross-biome predictors of bacterial diversity is questionable. Second, many studies have used short environmental gradients to predict bacterial alpha diversity, failing to identify non-linear (e.g. bimodal) relationships that are not apparent over short environmental gradients. For example, using a meta-analytical approach, Hendershot *et al.* (2017) suggested that the direction of the relationship between major environmental drivers (temperature and pH) and bacterial diversity are largely inconsistent at macroecological scales. Wide ranges of values in environmental factors, including high and low levels of temperature and pH, must be considered simultaneously in order to adequately identify the shape and direction of the relationship between environmental variables and alpha diversity. Finally, many previous studies have evaluated the role of single environmental drivers in the distribution of bacterial alpha diversity in isolation. It is clear, however, that multiple environmental predictors need to be considered simultaneously in order to adequately identify the global predictors of bacterial diversity. More importantly, the universal drivers of bacterial alpha diversity have never been assessed using data from multiple globally distributed terrestrial biomes, which have prevented us from identifying a unique set of environmental attributes that could be used to predict the distribution of bacterial diversity across and within globally distributed terrestrial biomes

Herein, we used the global dataset available online from Delgado-Baquerizo *et al.* (2018a) to produce a comprehensive catalogue of universally valid predictors of soil bacterial diversity (number of bacterial phylotypes) that apply within three widely-distributed biomes: arid, continental and temperate ecosystems, and across them. We used the machine learning algorithm Random Forest to provide a holistic view of the universal predictors of bacterial diversity. Finally, we used Structural Equation Modeling to achieve a system-level understanding of the relationships among major universal predictors and bacterial diversity. We expected soil pH to be a universal driver of bacterial diversity (Fierer and Jackson 2006; Lauber *et al.* 2009; but see Maestre *et al.* 2015; Hendershot *et al.* 2017). We aimed to move beyond the classical pH-bacterial diversity relationships, and identify further universal predictors of bacterial diversity that could be used

across widely different biomes. Also, using environmental information, we aim to generate a global atlas for bacterial diversity across the globe.

Material and Methods.

Field survey and soil sample collection

A detailed description of the dataset used in this paper is available from Delgado-Baquerizo *et al.* (2018a). This dataset has been used to identify key predictors and map the relative abundance of dominant bacterial taxa. The original dataset (hereafter 'Global Dataset') included 237 locations across a wide range of ecosystem types (forests, grasslands, and shrublands) with markedly contrasting vegetation, climate and soils. Soil sample collection took place between 2003 and 2015. The coordinates of each site were recorded *in situ* with a portable GPS, and the ecosystem type (grassland, shrubland, or forest) of each location recorded. At each site, a composite soil sample (uppermost 7.5 cm) was collected under the most common vegetation microsite present at a site. These microsites included trees, shrubs, grasses and open. After field collection, each soil sample was separated into two sub-samples, one which was immediately frozen at -20 ºC for molecular analyses, and the other air-dried for chemical analyses.

PCR-based 16S rRNA gene analyses

Soil DNA was extracted using the Powersoil® DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. A portion of the bacterial 16S rRNA gene (V3-V4 region) was sequenced using the Illumina MiSeq platform and the 341F/805R primer set. Bioinformatic processing was performed using a combination of QIIME (Caporaso *et al.* 2010), USEARCH (Edgar 2011) and UPARSE (Edgar *et al.* 2013). The resulting phylotype table was rarefied to 10000 sequences per sample. We further removed phylotypes (defined as operational taxonomic units at the 97% similarity) that were represented by only a single read across all samples. In addition, we removed any archaeal, chloroplast and mitochondrial phylotypes. The relative abundance of major phyla across biomes is shown in Fig. S1. Detailed information on the bacterial community composition of the Global dataset can be found in Delgado-Baquerizo *et al.* (2018a). We used bacterial richness (number of bacterial phylotypes) as our measure of bacterial diversity because it is the simplest measure of biodiversity, and is typically used to describe both belowground and aboveground organisms (Gotelli & Colwell 2001). However, bacterial richness

is well-known to be highly correlated with other alpha diversity indexes including Shannon diversity and Phylogenetic diversity (see Figs S7 and S8 in Delgado-Baquerizo *et al.* 2016a).

Environmental predictors

We measured soil pH, texture, total organic carbon (soil C), total nitrogen (soil N) and total phosphorus (soil P) concentrations using standard laboratory methods. Soil pH was measured in all soil samples with a pH meter, in a 1:2.5 mass: volume soil and water suspension. Texture (% of fine fractions: clay and silt) was determined according to Kettler *et al.* (2001). The concentration of soil total organic carbon (C) was determined using a wet chemistry method described in Anderson and Ingramm (1993). Soil total N was measured with a CN analyzer (Leco CHN628 Series, LECO Corporation, St Joseph, MI, USA) and total phosphorus (P) was measured using a SKALAR San++ Analyzer (Skalar, Breda, The Netherlands) after digestion with sulphuric acid.

We obtained information on maximum and minimum temperature, precipitation seasonality, and mean diurnal temperature range (MDR) for all sampling locations from the Worldclim database (www.worldclim.org), which has a 1 km resolution (Hijmans *et al.* 2005). In addition, for each site we estimated the Aridity Index (mean annual precipitation/potential evapotranspiration) from the Global Potential Evapotranspiration database (Zomer *et al.* 2008), which is based on interpolations provided by WorldClim (Hijmans *et al.* 2005). We used the Aridity Index rather than mean annual precipitation because aridity includes both mean annual precipitation and potential evapotranspiration, and is therefore a better measure of the long-term water availability at each site.

We used the Normalized Difference Vegetation Index (NDVI) as our proxy for net plant primary productivity (NPP, Pettorelli *et al.* 2005). This index provides a global measure of the "greenness" of vegetation across Earth's landscapes for a given composite period, and thus acts as a proxy of photosynthetic activity and large-scale vegetation distribution. NDVI data were obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Terra satellites (http://neo.sci.gsfc.nasa.gov/). We calculated the monthly average value for this variable between the periods of 2003 and 2015 (~10 km resolution), when all soil sampling was conducted. We obtained information on annual ultraviolet index (UV index) from the NASA's Aura satellite (https://neo.sci.gsfc.nasa.gov), which has a 50 km resolution. The UV index is a measure of the intensity of UV radiation ranging from 0 (minimal UV exposure risk) to 16 (extreme risk).

Statistical analyses within three major biomes

In this section, we focused on those locations for which we had sufficient information to model the distribution of bacterial alpha diversity within biomes. In particular, we focused on 223 locations across six continents (Fig. 1) that included three major global biomes: (1) arid (n = 102), (2) temperate (n = 81) and (3) continental (n = 40) ecosystems. Together, these biomes cover more than 70% of terrestrial ecosystems on Earth (excluding Antarctica). The biome grouping used in this study is based on the Köppen climate classification, one of the most widely used climate classification systems. We then conducted cross-biome analyses using the 237 locations included in the Global dataset, which also included tropical and polar ecosystems.

Random Forest analyses

We used Random Forest (Breiman, 2001) as described in Delgado-Baquerizo *et al.* (2016a) to identify the universal predictors of bacterial diversity across three major biomes: arid, continental and temperate. Our list of predictors included ultraviolet light (UV light), climate (Aridity Index, precipitation seasonality, maximum and minimum temperature, diurnal temperature range), soil properties (texture, soil C, N, P and pH), and vegetation types (presence/absence of forest or grassland). Shrublands were not available for all biomes, and were not explicitly included in our analyses. Random Forest is a novel machine-learning algorithm that extends standard classification and regression tree (CART) methods by creating a collection of classification trees with binary divisions. Unlike traditional CART analyses, the fit of each tree is assessed using randomly selected cases (33.3% of the data), which are withheld during its construction (out-of-bag or OOB cases). The importance of each predictor variable is determined by evaluating the reduction in prediction accuracy, i.e., increase in the mean square error between observations and OOB predictions, when the data for that predictor is randomly permuted. Random Forest is particularly recommended for datasets including categorical variables or variables with non-parametric distributions. Moreover, unlike multi-model inference using linear regressions or regression tree analyses, Random Forest alleviate any problems associated with multicollinearity in multivariate analyses by building bagged tree ensembles and including a random subset of features for each tree. All analyses were conducted using the rfPermute package (Archer *et al.* 2016) of the R statistical software (http://cran.r-project.org/).

Regression and ANOVA analyses

We used linear or quadratic relationships to evaluate the direction and shape of the relationship among universal environmental drivers and bacterial diversity as explained in Delgado-Baquerizo *et al.* (2016c). We used one-way ANOVAs to compare the diversity of bacteria across vegetation types: grasslands vs. forests for three biomes.

Structural Equation Modeling

After identifying the universal predictors of bacterial diversity within and across biome, we used Structural Equation Modeling (Grace 2006) to build a system-level understanding of the major direct and indirect effects of universal predictors of soil bacterial alpha diversity across the globe (*a priori* model available in Fig S2). Unlike regression or ANOVA, SEM allows us to separate multiple pathways of influence and view them as parts of a system. It is useful therefore for investigating the complex relationships among predictors commonly found in natural ecosystems (Grace 2006).

We included pH and UV light as polynomial variables in our SEM (see model selection for the Global dataset in Table S1). Consequently, these variables were included as composite variables made up of two components; pH and UV and pH^2 and UV^2 (Laliberté *et al.* 2014). The use of composite variables does not alter the underlying SEM model, but collapses the effects of multiple conceptually-related variables into a single composite effect, aiding interpretation of model results. With a good model fit, we were then free to interpret the path coefficients of the model and their associated P values. The probability that a path coefficient differs from zero was tested using bootstrap resampling. Bootstrapping is preferred to the classical maximum-likelihood estimation in these cases because in bootstrapping probability, assessments are not based on the assumption that the data match a particular theoretical distribution. The goodness of fit of SEM models was checked using the following: the Chi-square test, the root mean square error of approximation (RMSEA) and the Bollen-Stine bootstrap test (Schermelleh-Engel *et al.* 2003). SEM models were conducted with the software AMOS 20 (IBM SPSS Inc, Chicago, IL, USA).

Mapping of bacterial alpha diversity across the globe.

We used the prediction-oriented regression model Cubist (Quinlan 1993) as described in Delgado-Baquerizo *et al.* (2018a) to predict the distribution of bacterial alpha diversity across the globe. We included the following environmental predictors in our model: soil properties (soil C, soil pH and texture), climate (diurnal temperature range, maximum temperature, minimum temperature, Aridity Index and precipitation seasonality), net primary productivity, UV light and major

vegetation types (forests and grasslands). Online data is available for these variables via the ISRIC (global gridded soil information), Soil Grids (https://soilgrids.org/#!/?layer=geonode:taxnwrb_250m), the European Space Agency (http://due.esrin.esa.int/page_globcover.php), the WorldClim database (www.worldclim.org) and NASA satellites (https://neo.sci.gsfc.nasa.gov). We did not find high-resolution data for total N and total P, which were not included in this model. None of these variables were selected as universal predictors of bacterial diversity (see below). Global predictions of the distribution of major clusters were done on a 25 km resolution grid. We used the package Cubist in R to conduct these analyses (Kuhn *et al*. 2016). Finally, we cross-validated our map using two different approaches. First, we evaluated the correlation between observed and predicted data using our global dataset. Second, we used the database from the Earth Microbiome Project (Thompson et al. 2017), and evaluated the correlation between predicted bacterial diversity in our global map to that one independently obtained for 2004 soil samples in Thompson et al. (2017).

Results

Our Random Forest models explained 36, 43, 50 and 57% of the variation in the distribution of bacterial alpha diversity in arid, temperate and continental climates and in the whole Global dataset, respectively. These models provided evidence that only UV light, forest environments, soil C and pH are consistent predictors of bacterial diversity across three major biomes: arid, continental and temperate climates (Fig. 1b). These environmental variables were also important predictors of bacterial diversity across biomes (Fig. 1b). The relative importance of universal predictors of bacterial diversity shifted across biomes (Fig. 1b). Thus, soil C, soil pH and UV light were the most important universal predictors of bacterial diversity for temperate, continental and arid biomes, respectively (Fig. 1b).

Soil bacterial alpha diversity was lower in forests than grasslands (Figs. 2 and 3). Overall, UV light had a hump-shaped (or mostly negative) relationship with bacterial richness (Figs. 2 and 3). Bacterial diversity showed a positive or hump-shaped relationship with soil pH, and in general, was negatively related to soil C (Fig. 2). In arid ecosystems, soil C followed a hump-shaped relationship with bacterial diversity (<2%; Fig. 2G). In general, similar patterns (but see soil C) were observed when all biomes were analyzed together in the Global dataset (n = 237; Fig. 3),

which also included samples from polar and tropical biomes. Model selection is available in Table S1.

The Aridity Index was the most important climatic predictor in arid ecosystems, showing a positive relationship with bacterial alpha diversity (Fig. S3). Precipitation seasonality, which exhibited a hump-shaped relationship with bacterial diversity, was the most important climatic predictor in temperate ecosystems (Fig. S3). Finally, maximum temperature was the most important climatic predictor for continental climates, and showed a positive relationship with bacterial alpha diversity (Fig. S3).

Using only the four widespread predictors (UV light, soil C, soil pH and vegetation type; Fig. 4), our SEM was able to explain more than half (51%) of the variation in the distribution of bacterial diversity across the globe. Our *a priori* models attained an acceptable/good fit by all criteria in all cases, and thus no *post hoc* changes were made. Soil C had a direct negative effect on bacterial diversity (Fig. 4). Soil pH had a hump-shaped effect on alpha diversity (plotted in Fig. 3). Forest environments showed a direct negative effect on alpha diversity (Fig. 4). UV light had a direct negative (hump-shaped) effect on bacterial diversity (plotted in Fig. 3). Forest environments showed indirect negative effects on alpha diversity via reducing soil pH and increasing soil C (Fig. 4).

Finally, using environmental information, we generated a global map showing the distribution of soil bacterial alpha diversity (Fig. 5). Confirming our previous results, the cubist algorithm selected soil pH and C, UV light and vegetation type as important predictors for bacterial diversity. This model also selected Aridity Index, maximum temperature, diurnal temperature range, NPP and soil texture as important predictors. Our global map indicate that diversity of bacteria peak at global regions typically dominated by high pH, low C, grasslands and intermediate levels of radiation (Fig. 5; Fig. 6). Predicted and observed values for alpha diversity in this study were positively and significantly correlated (Fig. 5; Pearson´s r = 0.51; $P < 0.001$). Similarly, predicted and observed values for alpha diversity were positively and significantly correlated (Pearson´s r = 0.47; $P < 0.001$) using independent data from Thompson et al. (2017).

Discussion

Our work suggests that, from the wide range of selected environmental predictors included here, only UV light, forest environments, soil C and pH can be considered as significant universal

predictors of soil bacterial diversity valid within and across a wide range of biomes differing markedly in vegetation, climate and soil types. Our findings indicate that bacterial diversity is consistently lower in forests than grasslands, with low pH and relatively high carbon contents (>5%) across arid, temperate and continental biomes (Fig. 2 and 5). Soil pH is known to be one of the major environmental drivers of bacterial diversity at the global scale (Fierer and Jackson 2006), and therefore was expected to be a significant global driver of bacterial diversity across and within global biomes. Nonetheless, the importance of pH as a major predictor of diversity has been recently questioned (Maestre *et al.* 2015; Hendershot *et al.* 2017). Our study supports previous findings of an overall positive (or hump-shaped) relationship between soil pH and bacterial alpha diversity (Lauber *et al.* 2009), and provides strong support for the notion that soil pH is an important predictor of bacterial diversity worldwide. More importantly, our analyses provide strong evidence that UV light, soil C content and broad vegetation type (forest *cf.* grassland) should also be considered as universal predictors of the bacterial alpha diversity globally across major biomes. Together, the identified four universal predictors could predict more than half of the variation in the global distribution of bacterial diversity in our Random Forest and Structural Equation Models. Using environmental information, we generated a global map predicting the distribution of bacterial alpha diversity. Such a map reflects the results from the Random Forest and SEM models, using an independent Cubist model, and ultimately provides a novel atlas of the alpha diversity found in terrestrial ecosystems.

Ultraviolet light had an outstanding capacity to predict the distribution of bacterial diversity within and across biomes worldwide. Such a strong relationship between UV light and bacterial diversity had not been described previously for natural terrestrial ecosystems worldwide. In general (two out of three biomes, and when analyzing all data together), we found an overall hump-shaped relationship between UV light and diversity of bacteria. In other words, the diversity of bacteria peak at intermediate levels of UV light. High levels of UV light are known to reduce the abundance of bacteria of multiple species (Santos *et al.* 2012), supporting our results. More interestingly, an increase in UV light from very low to intermediate levels of intensity positively related to diversity of bacteria. Such a result is in agreement with the classic intermediate disturbance hypothesis (IDH; Wilkinson 1999), which suggest that species richness is maximized when ecological disturbance is neither too high nor too low. The IDH hypothesis suggest that intermediate levels of environmental disturbance support the coexistence of species capable of

surviving both low and high levels of disturbance, ultimately supporting the highest level of biodiversity. Our results provide novel evidence that UV light should be considered as a universal predictor for bacterial diversity. Another example of this relationships is the hump-shaped observed correlation between soil pH and bacterial diversity. A larger number of bacterial species can co-exist at intermediate neutral levels of pH.

Soil C content was negatively related to bacterial diversity within three major biomes and across all biomes. A strong negative correlation between soil C and bacterial diversity has been shown recently in temperate ecosystems from Scotland (Delgado-Baquerizo *et al.* 2017). Such a negative correlation might be explained by the strong competition to exclusion effect of microbial biomass on alpha diversity in locations with high levels of organic matter (Waldrop *et al.* 2006). Locations with higher soil C concentrations often support a greater abundance of soil bacteria and fungi, as supported by distributional maps of soil C stocks (Wieder *et al.* 2015) and microbial biomass (Fierer 2017). The only exception to this negative relationship occurs in arid zone soils with very low C contents (<2%). In arid, low carbon soils, bacterial diversity is positively correlated with soil C. Specifically, we found a hump-shaped relationship between soil C and bacterial diversity in arid environments, supporting the notion that soil C content limits the alpha diversity of bacterial communities in extremely low soil C environments (Fig. 2). This result has previously been reported in terrestrial ecosystems with very low C contents (e.g., arid and semiarid environments; Maestre *et al.* 2015; Neilson *et al.* 2017), or environments that included soil samples with extremely low level of soil C content (e.g. Antarctica; Delgado-Baquerizo *et al.* 2016a; 2018b).

Compared with grasslands, forest ecosystems have reduced levels of bacterial diversity across biomes (arid, continental and temperate ecosystems), and globally. Thus, ecosystem type (forest *vs*. grasslands) should also be considered as a global predictor of bacterial alpha diversity. This result accords with results from Crowther *et al.* (2014), who found an overall negative effect of forest environments *cf.* grasslands in North America. The reported negative relationship between forest environments and bacterial alpha diversity might be related to the indirect negative effects of forest environments on bacterial diversity via increasing soil C and reduced soil pH (Fig. 4), but also the strong competition of bacteria with plants and other soil organisms in rich and highly productive environments (Waldrop *et al.* 2006; Eldridge *et al.* 2017). Supporting this notion, Terrat *et al.* (2017; France) and Delgado-Baquerizo *et al.* (2018b; Australia) found that croplands

and grasslands had greater diversity of bacteria than forests. Our results suggest that grasslands, which are known to be a global hotspot of biodiversity (Roux *et al.* 2012), also have a higher level of alpha diversity than other terrestrial ecosystems, suggesting that our results are consistent with our understanding of how land use change (e.g. land clearing for farming) might affect microbial diversity. Conversion of forest to pasture is increasing markedly to meet global food demand, and is forecast to increase by 110% by 2050 (Tilman *et al.* 2011). Our study suggests that conversion of forest to grasslands will be accompanied by increases in bacterial alpha diversity. The importance of land use conversion as a major driver of bacterial alpha diversity has recently been highlighted by Szoboszlay *et al.* (2017) for terrestrial ecosystems across Europe, suggesting a positive effect of forest to pasture and cropland conversion on bacterial alpha diversity. Similarly, a previous study found an increase in the alpha diversity of particular groups of bacteria (e.g., *Verrucomicrobia*) following conversion from forest to pasture in Amazon (Ranjan *et al.* 2015).

No single climatic variable could predict the distribution of bacterial diversity across biomes. Our results suggest that climatic drivers of bacterial diversity cannot be considered globally applicable predictors of bacterial diversity, but that specific influences are limited to particular biomes. For example, we found that the Aridity Index, soil temperature and precipitation seasonality were the major climatic predictors of bacterial alpha diversity in arid, continental and temperate ecosystems, respectively (Fig. S3). These results accord with the results of studies by Maestre *et al.* (2015) and Zhou *et al.* (2016), who found that the Aridity Index and temperature were major predictors of bacterial alpha diversity in global drylands, and cold forest environments from North America, respectively. However, our results suggest that only precipitation seasonality and temperature were important climatic predictors of bacterial diversity in temperate and continental ecosystems, respectively. The importance of seasonality as a predictor of bacterial diversity was demonstrated by Delgado-Baquerizo *et al.* (2016b) using a meta-analysis approach.

Together, our results suggest that only a few environmental predictors including ecosystem type, UV light and soil C and pH, can be considered globally-applicable predictors of bacterial alpha diversity within and across biomes worldwide. This information can potentially be used to manage these soil organisms and the processes that they regulate in terrestrial ecosystems across large areas of the globe. Moreover, this environmental information can be used to predict the distribution of bacterial alpha diversity worldwide, which allowed us to generate a global atlas of bacterial diversity. Future studies could use our mapping approach to model the responses of soil

biodiversity under predicted global chance scenarios (e.g., warming and changes in precipitation regimes) at the global scale. Our results provide a holistic view of the direct and indirect effects of universal predictors on bacterial alpha diversity. This is likely to be an important and useful tool to help us understand potential changes in ecosystem functioning and the provision of essential services under global change scenarios.

Data accessibility: All data used in this study are publicly available in Figshare (https://figshare.com/s/82a2d3f5d38ace925492; DOI: 10.6084/m9.figshare.5611321).

Conflict of Interest

The authors declare no conflict of interest.

References

Bodelier, P.L.E. (2011). Toward understanding, managing, and protecting microbial ecosystems. *Front. Microb.*, 2, 80.

Breiman, L. (2001). *Random Forest Machine Learning*, 45, 5-32.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Method*s, 7, 335-336.

Cardinale, B.J., Emmett Duffy, J., Gonzalez, A., Hooper, D.U., Perrins, C., Venail, P. *et al.* (2012). Biodiversity loss and its impact on humanity. Nature 486, 59-67.

Crowther, T.W., Maynard, D.S., Leff, J.W., Oldfield, E.E., McCulley, R.L., Fierer, N. *et al.* (2014). Explaining the vulnerability of soil biodiversity to deforestation: a cross-biome study. *Glob. Chang. Biol.* 20, 2983-2994.

Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J. Encinar, D. *et al.* (2016a). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* 7, 10541.

Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Trivedi, P., Osanai, Y., Liu, Y-R., *et al.* (2016b) Carbon content and climate variability drive global soil bacterial diversity patterns. *Ecol. Monogr.* 86, 373-380.

Delgado-Baquerizo, M., Reich, P.B., Khachane, A.N., Campbell, C.D., Thomas, N., Freitag, T., *et al.* (2017). It is elemental: soil nutrient stoichiometry drives bacterial diversity. *Environ. Microbiol.* 19, 1176–1188.

ado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D. *et al.* (2018a) A global atlas of the dominant bacteria found in soil. *Science* 19, 320-325.

Delgado-Baquerizo M, Reith F, Dennis PG, Hamonts K, Powell JR, Young A, Singh BK, Bissett A (2018b). Ecological drivers of soil microbial diversity and soil biological networks in the Southern Hemisphere. Ecology (in press). doi: 10.1002/ecy.2137.

Delgado-Baquerizo, M., Giaramida, L., Reich, P.B., Khachane, A.N., Hamonts, K., Edwards, C., Lawton, L., Singh, B.K. (2016c). Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. *J. Ecol.* 104, 936–946.

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B. *et al.* (2014). Defaunation in the Anthropocene. *Science* 345, 401-406.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194-2200.

Eldridge, D.J., Delgado-Baquerizo, M., Travers, S.K., Val, J., Oliver, I., Hamonts, K. *et al.* (2017). Competition drives the response of soil microbial diversity to increased grazing by vertebrate herbivores. *Ecol*., 98, 1922-1931.

Fierer, N., Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA,* 103, 626–631.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L. *et al.* (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci. USA*, 26, 21390-21295.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Micro.,* 15, 579-590.

Gotelli, N.J., Colwell, R.K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.,* 4, 379-391.

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

Hendershot, J.N., Read, Q.D., Henning, J.A., Sanders, N.J., Classen, A.T. (2017). Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. *Ecol.*, 98, 1757-1763.

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. *Intern. J. Climat.*, 25, 1965-1978.

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S. *et al.* (2005). Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monog.*, 75, 3-35.

Jing, X., Sanders, N.J., Shi, Y., Chu, H., Classen, A.T., Zhao, K. *et al.* (2015). The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. *Nat. Commun.*, 6, 8159.

Laliberté, E., Zemunik, G., Turner, B.L. (2014). Environmental filtering explains variation in plant diversity along resource gradients. *Science*, 345, 1602-1605.

Lauber, C.L., Hamady, M., Knight, R., Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.*, 75, 5111-5120.

Maestre, F.T., Delgado-Baquerizo, M., Jeffries, T.C., Eldridge, D.J., Ochoa, V., Gozalo, B. *et al.* (2015). Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proc. Natl. Acad. Sci. USA*, 112, 15684–15689.

Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.M., Tucker, C.J., Stenseth, N.C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol*. 20, 503-510.

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C.M. *et al.* (2013). Loss in microbial diversity affects nitrogen cycling in soil. *ISME J.* 7, 1609-1619.

Prober, S.M., Leff, J.W., Bates, S.T., Borer, E.T., Firn, J., Harpole, W.S. *et al.* (2015). Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecol. Lett.*, 18, 85-95.

Ramirez, K.S., Leff, J.W., Barberán, A., Bates, S.T., Betley, J., Crowther, T.W. (2014). Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. B.*, 281, 20141988

Ranjan, K., Paula, F.S., Mueller, R.C., Jesus, E.C., Cenciani, K., Bohannan, B.J.M. *et al.* (2015). Forest-to-pasture conversion increases the diversity of the phylum Verrucomicrobia in Amazon rainforest soils. *Front. Microbiol.*, 6, 779.

Rousk, J., Brookes, P.C., Baath, E. (2010). The microbial PLFA composition as affected by pH in an arable soil. *Soil Biol. Biochem*., 42, 516-520.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S. *et al.* (2012). Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS One*, 7, e33641.

Szoboszlay, M., Dohrmann, A.B., Poeplau, C., Don, A., Tebbe, C.C. (2017). Impact of land-use change and soil organic carbon quality on microbial diversity in soils across Europe. *FEMS Microbiol. Ecol*. 1, 93.

Terrat, S., Horrigue, W., Dequietd, S., Saby, N.P.A., Lelièvre, M., Nowak, V. *et al.* (2017). Mapping and predictive variations of soil bacterial richness across France. *PLoS ONE* 12, e0186766.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J. *et al.* (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 23, 457-463.

Tilman, D., Balzer, C., Hill, J., Befort, B.L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA*, 13, 20260-20264.

Waldrop, M.P., Zak, D.R., Blackwood, C.B., Curtis, C.D., Tilman, D. (2006). Resource availability controls fungal diversity across a plant diversity gradient. *Ecol. Lett.*, 9, 1127-1135.

Whitford, W.G. (1978). Foraging in seed-harvester ants *Pogonomyrmex* spp. *Ecol*., 59, 185–189.

Wieder, W.R., Bonan, G.B., Allison, S.D. (2013). Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Change*, 3, 909–912.

Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D. *et al.* (2016). Temperature mediates continental-scale diversity of microbes in forest soils. *Nat. Comm*., 7, 12083.

Zomer, R.J., Trabucco, A., Bossio, D.A., Verchot, L.V. (2008). Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric. Ecosyst. Envir*., 126, 67-80.

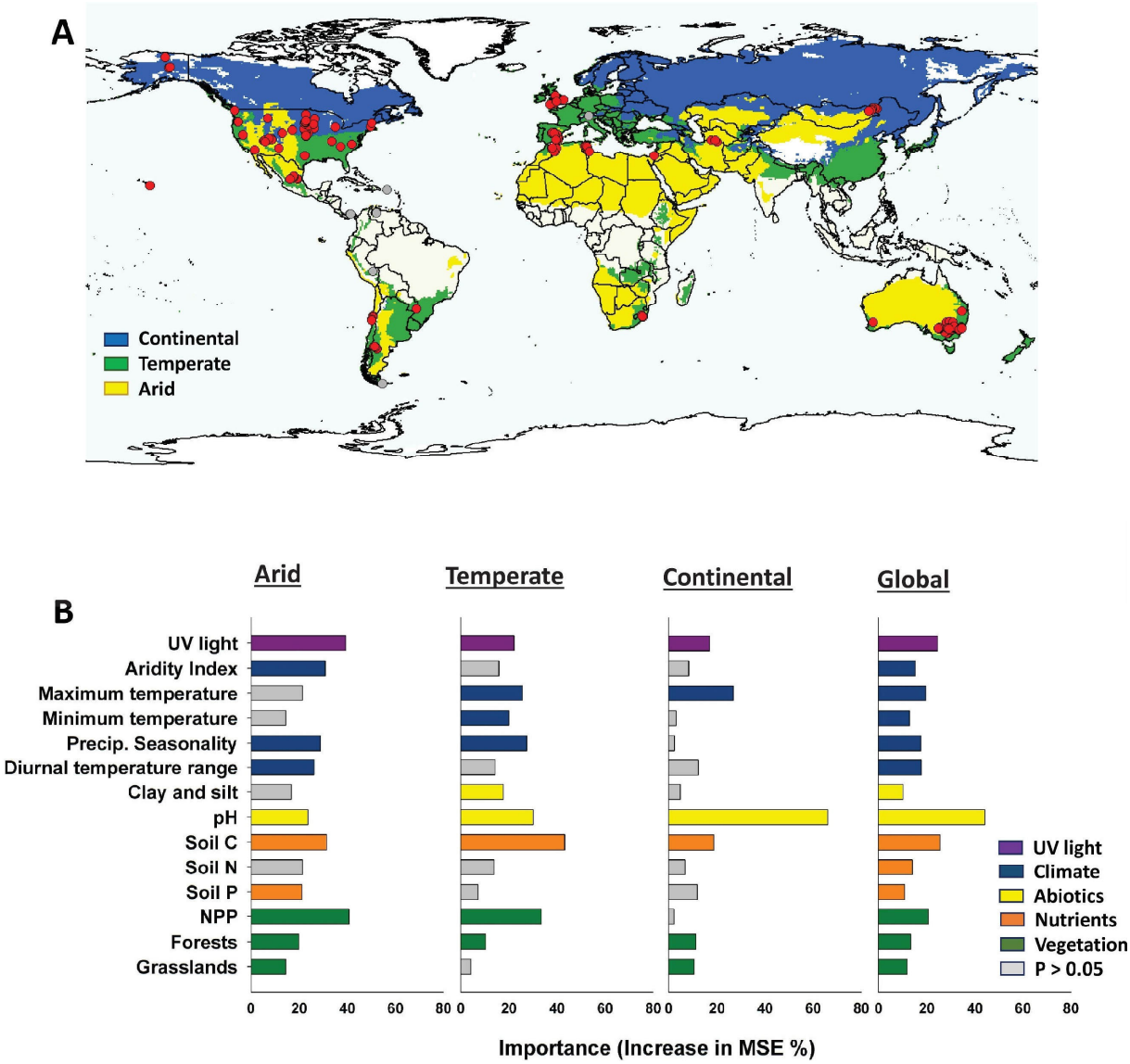


Figure 1. Location for the 237 sites globally distributed across arid, continental and temperate climates (red circles) and others (grey circles) (a). Results from a Random Forest aiming to identify the main significant ($P < 0.05$) environmental predictors of bacterial alpha diversity (b). MSE = Mean Square Error.

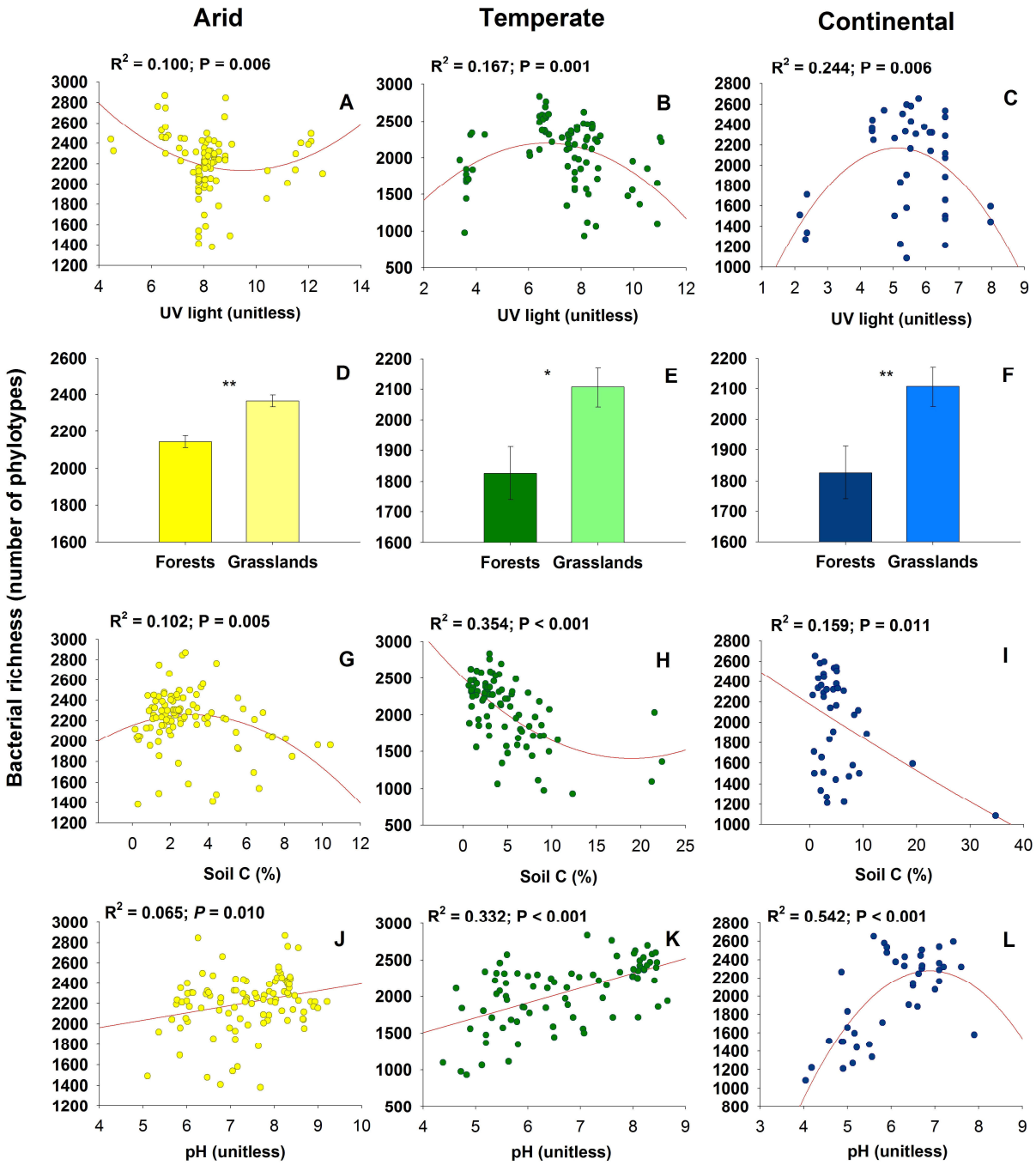


Figure 2. Relationship between universal environmental predictors and alpha diversity of bacterial communities across three globally distributes biomes.

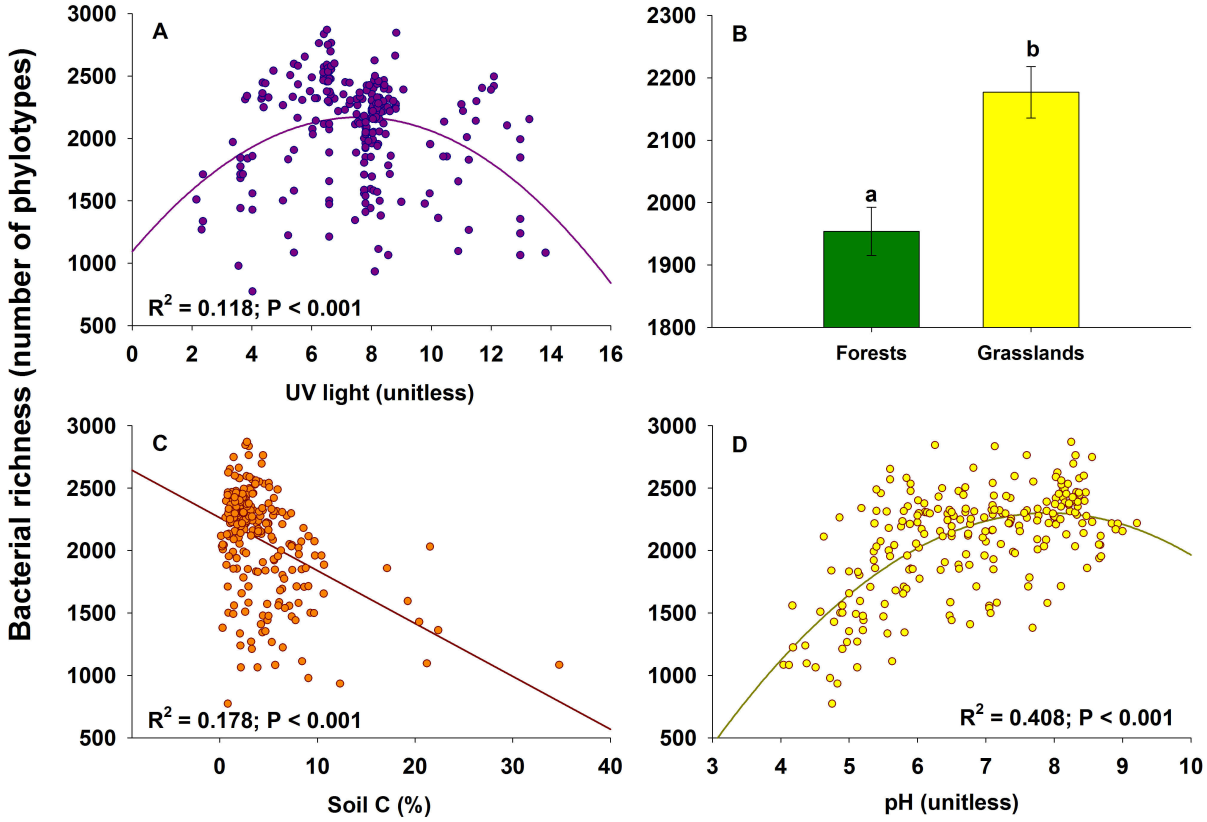


Figure 3. Relationship between universal predictors and bacterial alpha diversity across the globe.

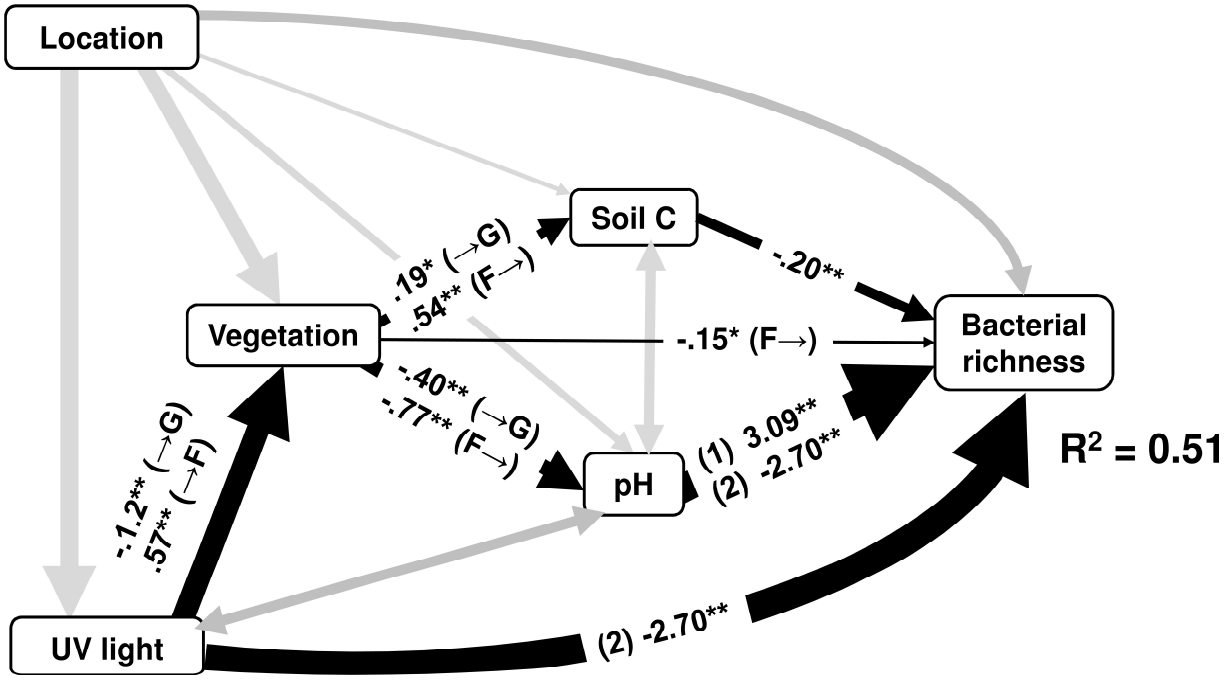


Figure 4. Structural equation model describing the selected direct and indirect effects of the universal predictors on bacterial alpha diversity. Numbers adjacent to arrows are indicative of the effect size of the relationship. R^2 denotes the proportion of variance explained. Significance levels of each predictor are *P<0.05, **P<0.01. F = forests, G = grasslands. There was a non-significant deviation of the data from the model (χ^2 = 0.22, df = 1; P=0.63; RMSEA P = 0.73; Bootstrap P = 0.68). Other significant effects, not included in this graph for simplicity (e.g., those from location), are included in Table S2. The numbers (1) and (2) superimposed on the arrows in our model indicate the coefficients coming from the two component of a quadratic regression. F = forest; G = Grasslands.

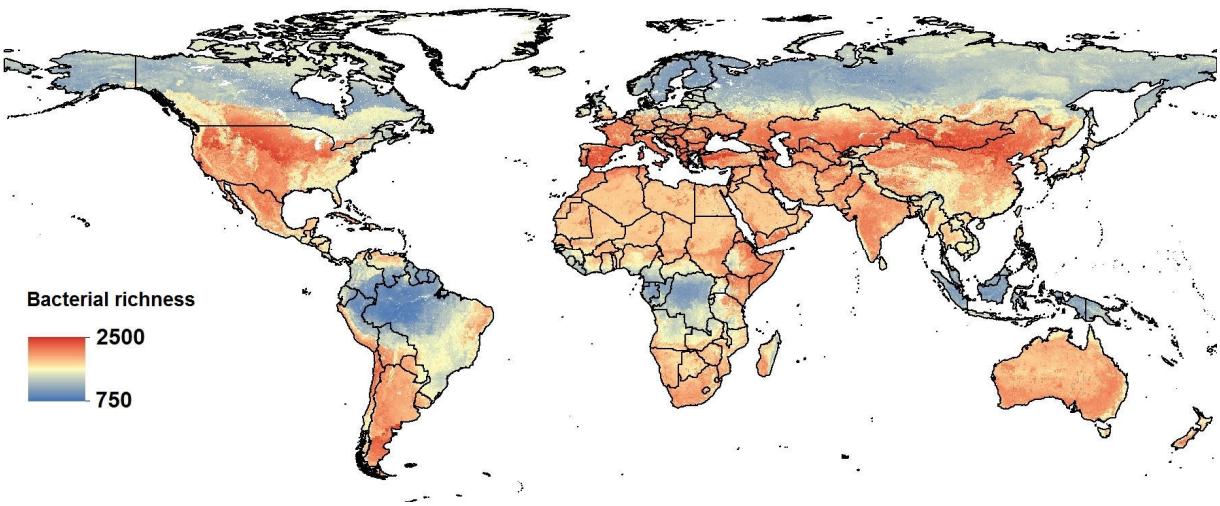
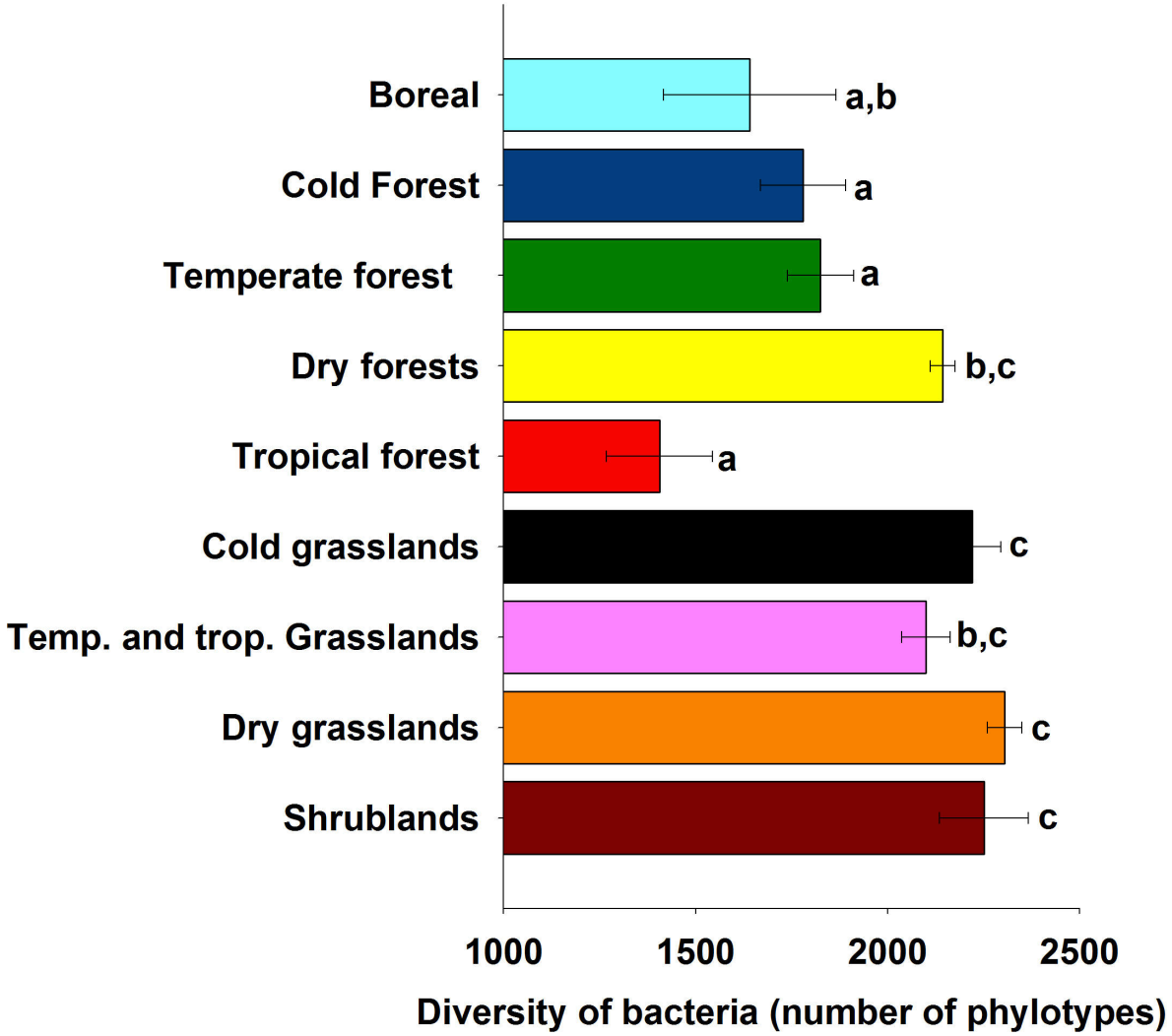


Figure 5. Predicted global distribution of bacterial alpha diversity across the globe.



592

593 **Figure 6**. Mean (± SE) bacterial richness across major ecosystem types (n = 237). Ecosystem type
594 classification followed the Köppen climate classification and the major vegetation types found in
595 our database. Grasslands include both tropical and temperate grasslands. Shrublands include polar,
596 temperate and tropical shrublands. Letters indicate post-hoc analyses from PERMANOVA.

597

598

599