Carbon content and climate variability drive global soil bacterial diversity patterns

Manuel Delgado-Baquerizo,1,6 Fernando T. Maestre,2 Peter B. Reich,1,3 Pankaj Trivedi,1 Yui Osanal,1 Yu-Rong Liu,1,4 Kelly Hamonts,1 Thomas C. Jeffries,1 and Brajesh K. Singh1,5

1Hawkesbury Institute for the Environment, Western Sydney University, Penrith, 2751, New South Wales Australia
2Área de Biodiversidad y Conservación, Departamento de Biología y Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Calle Tulipán Sin Número, 28933, Móstoles Spain
3Department of Forest Resources, University of Minnesota, St. Paul, Minnesota 55108 USA
4State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, 100085 China
5Global Centre for Land-Based Innovation, Western Sydney University, Penrith, 2751, New South Wales Australia

Abstract. Despite the vital role of microorganisms for ecosystem functioning and human welfare, our understanding of their global diversity and biogeographical patterns lags significantly behind that of plants and animals. We conducted a meta-analysis including ~600 soil samples from all continents to evaluate the biogeographical patterns and drivers of bacterial diversity in terrestrial ecosystems at the global scale. Similar to what has been found with plants and animals, the diversity of soil bacteria in the Southern Hemisphere decreased from the equator to Antarctica. However, soil bacteria showed similar levels of diversity across the Northern Hemisphere. The composition of bacterial communities followed dissimilar patterns between hemispheres, as the Southern and Northern Hemispheres were dominated by Actinobacteria and Acidobacteria, respectively. However, Proteobacteria was co-dominant in both hemispheres. Moreover, we found a decrease in soil bacterial diversity with altitude. Climatic features (e.g., high diurnal temperature range and low temperature) were correlated with the lower diversity found at high elevations, but geographical gradients in soil total carbon and species turnover were important drivers of the observed latitudinal patterns. We thus found both parallels and differences in the biogeographical patterns of aboveground vs. soil bacterial diversity. Our findings support previous studies that highlighted soil pH, spatial influence, and organic matter as important drivers of bacterial diversity and composition. Furthermore, our results provide a novel integrative view of how climate and soil factors influence soil bacterial diversity at the global scale, which is critical to improve ecosystem and earth system simulation models and for formulating sustainable ecosystem management and conservation policies.

Key words: bacterial composition; biodiversity; diurnal temperature range; global biogeography; soil carbon; terrestrial ecosystems.

Introduction

Soil microbes are among the most abundant and diverse organisms on Earth, and are responsible for key ecosystem functions and services, including decomposition, nutrient cycling and climate regulation (Singh et al. 2009, Bodelier 2011, Bardgett and van der Putten 2014, Tedersoo et al. 2014, Delgado-Baquerizo et al. 2016). However, there are important gaps in our understanding of the patterns and drivers of soil microbial diversity at a global scale (Bodelier 2011, Bardgett and van der Putten 2014). These knowledge gaps have kept microbial diversity outside of ongoing debates about global biodiversity loss, conservation and sustainable management policies (Bodelier 2011, Bardgett and van der Putten 2014) and have precluded the inclusion of microbial communities in global biogeochemical models (Bodelier 2011, Wieder et al. 2014). Understanding soil microbial diversity patterns is thus critical to establish effective policies to preserve microbial diversity hotspots and key ecosystem functions and services provided by soil microbes (Singh et al. 2009, Bodelier 2011, Bardgett and van der Putten 2014, Tedersoo et al. 2014).

One of the main reasons for the current knowledge gaps is that we lack a strong theoretical foundation to explain microbial diversity and structural patterns in terrestrial ecosystems. This is surprising though, as there is a large body of potentially relevant ecological theory on plant and animal communities from which to draw conclusions on the biogeography of soil microbes. For instance, the increase in plant and animal diversity from the poles to the tropics constitute a major global biogeographical pattern in terrestrial ecosystems (MacArthur 1975, Rohde 1992, Begon et al. 2006). Although altitudinal patterns in plant and animal diversity are more controversial (Rahbek 2005), a thorough review of the literature suggests that plant and animal richness often
follow a hump-shaped relationship with altitude in terrestrial ecosystems, with species richness typically decreasing from mid to high elevations (MacArthur 1975, Rohde 1992, Rahbek 2005, Begon et al. 2006, Sundaqvest et al. 2013). However, whether these patterns apply to soil microbial diversity is largely unknown (Singh et al. 2009, Bardgett and van der Putten 2014). The extreme climatic conditions (e.g., low temperature and high seasonality, Begon et al. 2006), relatively short time since glaciations (MacArthur 1975, Rohde 1992, Begon et al. 2006), and low net primary productivity (NPP) found at mid-to-high latitudes may explain the low number of plant and animal species found in these areas (MacArthur 1975, Rohde 1992, Rahbek 2005, Begon et al. 2006). While the climate harshness hypothesis (Begon et al. 2006, Young et al. 2013) may directly apply to microbial diversity at the top of high mountains, the different energy sources used by micro- and macro-organisms (soil organic matter vs. solar or NPP; MacArthur 1975, Rohde 1992, Begon et al. 2006, Allison et al. 2010, Trivedi et al. 2013) and their capacity to migrate toward new environments (low vs. very high; Szekely et al. 2013) may promote divergence in global biodiversity patterns (MacArthur 1975, Rohde 1992, Currie et al. 2004, Begon et al. 2006). For a wide variety of hypothesized reasons, including their large body size and maintenance costs compared to microbes, the diversity of plants and animals may peak where NPP is highest, which is in the tropics (MacArthur 1975, Rohde 1992, Begon et al. 2006). However, most soil microbial communities rely on processes such as organic matter decomposition to obtain energy (Allison et al. 2010, Trivedi et al. 2013), and thus may peak in areas with high soil organic matter content, which do not necessarily correspond to those with the highest plant productivity (Xu et al. 2013, Wieder et al. 2014).

The equator divides the global climate (e.g., annual mean temperature) and terrestrial net primary productivity patterns roughly into two mirror images; accordingly, plant and animal diversity both decrease from the equator to the poles (Rohde 1992, Begon et al. 2006). However, recent studies suggest that soil total carbon (C) and microbial biomass C decrease from northern high-latitude areas to the Southern Hemisphere (Xu et al. 2013, Wieder et al. 2014). The high soil total C found in northern high latitudes is likely the result of organic matter accumulation due to slow decomposition in taiga and tundra ecosystems, despite extreme environmental conditions strongly limiting NPP and plant and animal diversity in these areas (Ernakovich et al. 2014). Given the characteristics of microbes, such as their rapid generational time compared to plants and animals (Schmidt et al. 2007) and their ability to survive abiotic stress in dormant forms without losing viability (Jones and Lennon 2010), our central hypothesis is that microbial diversity will peak in northern high latitude regions and decrease continuously towards high latitudes in the Southern Hemisphere, mirroring soil total C patterns (Xu et al. 2013, Wieder et al. 2014).

Previous studies have highlighted the importance of soil properties such as pH (Fierer and Jackson 2006, Fierer et al. 2009, Lauber et al. 2009), spatial influence (Horner-Devine et al. 2004, Martiny et al. 2011, Hanson et al. 2012) and soil organic matter (Siciliano 2014, Maestre et al. 2015) as drivers of variation in soil microbial diversity. However, the local or regional spatial scales at which these studies have been conducted (Horner-Devine et al. 2004, Fierer and Jackson 2006, Fierer et al. 2009, Lauber et al. 2009, Steven et al. 2013), and the low taxonomic resolution of the techniques used to estimate microbial diversity limit their ability to accurately assess its patterns and drivers at the global scale (Horner-Devine et al. 2004, Fierer and Jackson 2006, Fierer et al. 2009).

Here, we evaluated whether soil bacterial diversity peaks where organic matter (i.e., soil total C) accumulation is high (Xu et al. 2013, Wieder et al. 2014), following our central hypothesis, or alternatively follows similar biogeographical patterns as those found for vascular plants and animals. We also assessed the relative importance of multiple climate and soil variables as drivers of this diversity. To do this, we conducted a meta-analysis of available data on soil bacterial diversity and composition evaluated with next generation sequencing techniques (454 Pyrosequencing and Illumina MiSeq), which included locations from all continents and biomes (Fig. 1; Appendix S1). We collected data on bacterial diversity, as measured by the Shannon diversity index of operational taxonomic units (OTUs) at 97% of similarity (bacterial diversity; n = 390 soil samples), and composition (10 major phyla of bacteria; n = 588 soil samples) from the literature (Appendix S1) and unpublished data. Moreover, we collected information on different spatial (latitude and altitude), climatic (annual mean temperature, precipitation seasonality, aridity, and diurnal temperature range) and soil (total soil C, C:N ratio, and pH) variables associated with the sample locations, and conducted multiple in silico and statistical analyses to ensure that the use of different primer sets, soil depths and sequencing sampling efforts among the studies included in our database did not bias the results obtained. Altogether, our study aims to provide a review of the current literature and a conceptual framework combining ecological theory and novel statistical tools to help future studies addressing global patterns in microbial diversity in terrestrial ecosystems.

**Materials and Methods**

**Data collection**

Data on soil bacterial diversity based on next generation sequencing techniques (either 454 or MiSeq) were collected from both published and unpublished data. A search using SCOPUS during September 2014 used the following keyword combinations: (1) “bacterial community” AND “soil” AND “pyrosequencing”; and (2) “bacterial community” AND “soil” AND “Illumina”. Approximately 300 references were found. Within these references, studies were included in our analyses if they met the following criteria: (1) they were carried out in the field in terrestrial ecosystems; (2) they contained the
spatial location where they were carried out (latitude and longitude), as well as data on soil total C and pH; (3) they provided information on bacterial diversity at 97% of similarity; (4) they included data on the relative abundance of soil bacterial phyla; and (5) they used next generation sequencing techniques (either 454 or Miseq). From those experimental studies that manipulated environmental conditions (e.g., nutrients or climatic conditions) we only used data from the control treatment.

A total of 102 articles containing data either on bacterial diversity or composition were obtained from our literature search (see Appendix S1 for complete list of studies). We found 546 soil samples from 397 sites. We completed this database with data from 38 sites (58 soil samples) forming part of the National Ecological Observatory Network (NEON; 24 soil samples) in the United States, with 12 soil samples from the Canadian MetaMicroBiome Library and with unpublished data from a site from England (four soil samples), three sites from Australia (12 soil samples), and three sites from Scotland (six soil samples; unpublished data). This increased the total number of samples used to 604 independent soil samples from ~435 sites (Fig. 1). Bacterial Shannon diversity and composition were available for 65% and 97% of these samples (390 and 588 samples), respectively.

For each soil sample included in the database, the following information was included:

1) Geographic coordinates: latitude and longitude.
2) Altitude: this variable was obtained from digital models using the Worldclim database (available online; Hijmans et al. 2005). Altitude gaps were completed in the data set using Google Earth.
3) Soil properties: soil total C, C:N ratio, and pH. In the data set, 95% of the cases used either a dry combustion method with an elemental analyzer (~71%) or the Walkley and Black (1934) dichromate oxidation (~24%) method to determine soil C. Similarly, 95.4% of the pH measurements in the studies were performed using a water–soil mix and a pH meter. The amount of soil C was transformed into mass percent (%) prior to analyses.
4) Climate: mean annual temperature and aridity index (mean annual precipitation/potential evapotranspiration), as well as daily temperature range and rainfall seasonality were determined from the Worldclim database (Hijmans et al. 2005) and Zomer et al. (2006). Climate gaps in the data set were completed using local and regional databases. For clarity, we used aridity (maximum value of aridity index

\[ \text{Aridity index} = \frac{\text{Mean annual precipitation}}{\text{Potential evapotranspiration}} \]

Fig. 1. Locations of the sites included in this study. The sites used for the analyses of soil bacterial diversity are in white (390 soil samples), while both white and black dots indicate the sites included in the analyses of soil bacterial composition (588 soil samples).
Most studies extracted DNA using largely standardized DNA extraction kits (e.g., PowerSoil of MOBIO Laboratories, Carlsbad, California, USA). Ninety-seven percent of the data used were analyzed using a 454 pyrosequencing sequencer (454 life-sciences, Branford, Connecticut, United States; Branford, Connecticut, United States), while the remaining part was analyzed using a MiSeq Illumina sequencer (Illumina, San Diego, California, United States). All the data compiled here were published/obtained between 2005 and 2014. Furthermore, 31%, 18%, and 9% of the soil samples included were sequenced using regions V4 (i.e., 515F to 816R), V1–V3 (i.e., 8F to 556R), and V3–V4 (i.e., 338F to 907R), respectively. The rest of the soil samples were sequenced using either different combinations of these primer sets (between 8F-816R, V1–V4; ~31%) or other regions (V4–V8; ~11%). Thirty-nine percent and 24% of the soil samples were analyzed with QIIME (Caporaso et al. 2010) and MOTHUR (available online), respectively. The remaining samples were either analyzed with a custom-made pipeline or the information was not provided (5%). Taxonomy assignments were made in 58%, 15%, and 13% of samples using RDP, GreenGenes (DeSantis et al. 2006), and SILVA (Quast et al. 2013) databases, respectively. The data used in this study included sites from all continents (Fig. 1), and contained wide gradients of climatic conditions and soil properties, as well as multiple management regimes and biomes. For example, altitude, annual mean temperature, and precipitation ranged from 0 to 4600 m, from −19.8° to 27.5°C, and from 8 to 4378 mm, respectively. Soil properties also showed wide gradients in pH (3.1–10.4) and soil total C (0.02–46.9%). In addition, soil samples belonged to multiple ecosystem types. The different biomes included in our database are (proportion for both the Northern and Southern Hemispheres): agro-systems (29% and 22%), boreal and hemi-boreal forests (18% and 0%), temperate and continental (16% and 8%), drylands (12% and 10%), temperate forests (14% and 14%), tropical forests (4% and 14%), and tundra (5% and 32%).

Shannon diversity (Shannon 1948) was selected as our metric for alpha diversity because it is highly recommended when analyzing microbial diversity (Haegeman et al. 2013). It has been shown to reduce the bias in relation to other diversity metrics, such as the number of OTUs, when comparing data from multiple sources (He et al. 2013). Since the comparison among studies may not be straightforward due to the manner in which the Shannon index is calculated, we took the following precautions to minimize the noise among studies: (1) Shannon diversity was used based on OTUs built at 97% sequence similarity. This level of similarity was used because it has been widely employed for bacterial diversity across different studies (see Appendix S1 for examples); (2) values from this index were used based on natural logs (e.g., Shannon indices were multiplied by ln2 to allow comparison with non-QIIME-based studies; Schneider et al. 2013, Carr et al. 2013); (3) when available, information was collected on the rarefaction level applied prior to calculating the Shannon index; (4) information on the soil depth in which each sample was collected; and (5) information was recorded on the primer set and region used in the next generation sequencing analyses. The former three considerations were explicitly taken into account in our analyses as explained in the next sections.

Statistical analyses I. Exploring geographical patterns and evaluating major drivers of microbial diversity and composition

Latitudinal and altitudinal patterns in soil bacterial diversity were explored by fitting linear regressions between these geographical variables and the Shannon index measured as described in Data collection. Latitude was used instead of the distance from the equator (i.e., absolute latitude) based on preliminary analyses that showed different latitudinal patterns for bacterial diversity in the Northern and Southern Hemispheres. Altitude was log_{10}-transformed to improve normality. P values were nonparametrically tested using the distlm function from PERMANOVA (McArdle and Anderson 2001). In addition, differences in soil bacterial diversity and composition (main bacterial phyla) among different geographical regions (Arctic Circle [>66° N], middle latitude north [23–66° N], equator [23° N to 23° S], middle latitude south [23–66° S] and Antarctic Circle [>66° S]; Marsh and Kaufman 2013) were checked independently using one-way PERMANOVA, with geographical region as a fixed factor in both cases (Anderson 2001). Additionally, relationships between the relative abundance of a selected range of main bacterial phyla and latitude, distance from the equator, altitude, climate, and soil properties were explored using Spearman’s rank correlations.

All the analyses explained above were performed at the level of soil samples (n = 390 and 588 for soil bacterial diversity and composition, respectively). Analyses at the soil sample level were conducted to capture the widest range of variation and resolution possible in soil properties (i.e., soil C, C:N, and pH). These predictors, which have a huge influence on bacterial diversity and composition in terrestrial ecosystems (Fierer and Jackson 2006, Lauber et al. 2009, Siciliano 2014, Tedersoo et al. 2014, Maestre et al. 2015), are highly variable at a local scale (Cambardella et al. 1994, Gallardo 2003, Gallardo and Parama...
A mandatory question was whether any spatial pattern observed at a sample-level scale is maintained at a site-level scale. To account for this important point, we explored the relationship between latitude and altitude with bacterial diversity at a site level \( (n = 278) \). To calculate site-level estimates of bacterial diversity and composition, we used the averaged values of all soil samples from the same site, keeping only a single value per site.

A classification Random Forest analysis (Breiman 2001) was performed to assess which variables (spatial, climate, and soil properties) were the most important drivers of the variation found in bacterial diversity (Shannon index, our dependent variable). 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 (Wei et al. 2010). Unlike traditional CART analyses, the fit of each tree is assessed using randomly selected cases (one-third 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 decrease in prediction accuracy (i.e., increase in the mean square error between observations and OOB predictions) when the data for that predictor are randomly permuted. This decrease is averaged over all trees to produce the final measure of importance (Wei et al. 2010). This accuracy importance measure was computed for each tree and averaged over the forest (5000 trees). These analyses were conducted using the randomForest package (Breiman 2001) for the R statistical software, version 3.0.2 (R Development Team 2013). The significances of the model and the cross-validated \( R^2 \) were assessed with 5000 permutations of the response variable using the A3 R package (Fortmann-Roe 2013). Similarly, the significance of the importance measures of each predictor (here soil variables) on the response variable (Shannon diversity) was assessed by using the rffPermute package for R (Liau and Wiener 2002).

Once Random Forest analyses were completed, structural equation modelling (SEM; Grace 2006) was used to evaluate the relationships between geographical patterns (latitude and altitude), climate (annual mean temperature, aridity, diurnal temperature range, and rainfall seasonality), soil properties (pH, soil total C, and C:N ratio) and soil bacterial diversity (Shannon index). The first step in SEM requires establishing an a priori model based on the known effects and relationships among the drivers of macro and microorganisms diversity (Fig. S1 in Appendix S1). Some data manipulation was required prior to modelling. The bivariate relationships between all variables were checked to ensure that a linear model was appropriate. Some curvilinear relationships were identified among our variables. Several variables showed a curvilinear relationship with latitude, such that areas more close to the equator tended to be different from areas farther from the equator. This was handled by expressing latitude as distance from the equator (i.e., \( \text{latitude}^2 \)). However, because other variables, such as soil total C and bacterial diversity, were driven by latitude itself, both latitude and distance from the equator were included in the model as composite variables (see details on composite variables that follow). In addition, soil bacterial diversity is largely known to be curvilinearly influenced by pH, and this relationship was well described by a second-order polynomial. To introduce polynomial relationships into our model, the square of soil pH was calculated and introduced into the model using the composite variable approach. The distribution of all of our endogenous variables was examined, and their normality was tested. Soil total C, C:N ratio and altitude were log-transformed to improve normality. Moreover, because of the low importance of C:N ratio as a driver of bacterial diversity, suggested by the Random Forest analyses, the direct effect of C:N ratio on bacterial diversity was removed to get a degree of freedom that allowed the testing for goodness of fit on the models. However, soil total C and C:N ratio were allowed to covariate in our SEM, maintaining the ability to test the indirect effect of C:N ratio on bacterial diversity via soil total C.

After attaining a satisfactory model fit, composite variables were introduced into the model. 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 with the interpretation of model results (Grace 2006). Climatic variables (annual mean temperature, aridity, diurnal temperature range, and precipitation seasonality), geographical coordinates (latitude and distance from the equator), and soil pH (pH and \( \text{pH}^2 \)) were included as composite variables in our model. In the SEM, soil total C and C:N ratio were grouped as organic matter for graphical simplicity but not combined in analyses. With a good model fit and composite variables constructed, we were free to interpret the path coefficients of the model and their associated \( P \) values. A path coefficient is analogous to the partial correlation coefficient, and describes the strength and sign of the relationship between two variables (Grace 2006). Since some of the variables introduced were not normally distributed, the probability that a path coefficient differs from zero was tested using bootstrap (Grace 2006). Bootstrapping is preferred to the classical maximum-likelihood estimation in these cases because bootstrapping probability assessments are not based on the assumption that the data match a particular theoretical distribution. Thus, data are randomly sampled with replacement to arrive at estimates of standard errors that are empirically associated with the distribution of the data found in the samples (Grace 2006). When these data manipulations were completed, the model was parameterized using our data set and tested for its overall goodness of fit. There is no single universally accepted test of overall goodness of fit for SEM, applicable in all situations regardless of sample size or data distribution. The chi-square test \( (\chi^2) \) of the model has a good fit when \( 0 \leq \chi^2 \leq 2 \) and \( 0.05 < P \leq 1.00 \); Schermelleh-Engel et al. 2003) and the root mean square error of approximation (RMSEA; the model has a good fit when \( \text{RMSEA} 0 \leq \text{RMSEA} \leq 0.05 \) and \( 0.10 < P \leq 1.00 \);
Statistical analyses II: Taxonomic community similarity and bacterial composition across different regions and microbial turnover

The taxonomic community similarity within different latitudinal regions (Arctic Circle \([>66^\circ\text{ N}]\), middle latitude north \([23–66^\circ\text{ N}]\), equator \([23^\circ\text{ N} to 23^\circ\text{ S}]\), middle latitude south \([23–66^\circ\text{ S}]\), and Antarctic Circle \([>66^\circ\text{ S}]\)) was calculated based on the most common bacterial phyla (relative abundance of Proteobacteria, Acidobacteria, Actinobacteria, Verrucomicrobia, Bacteroidetes, Firmicutes, Gemmatimonadetes, Planctomycetes, and others). Differences among latitudinal ranges were evaluated using pairwise post-hoc tests in PERMANOVA (Anderson 2001). Microbial composition data were log-ratio-transformed before these analyses to reduce interdependence. PERMANOVA analysis and taxonomic community similarity were conducted using the Bray–Curtis distance with the PERMANOVA+ add-on of the PRIMER statistical package (PRIMER-E Ltd., Plymouth Marine Laboratory, Plymouth, UK). These analyses were conducted separately at the sample and site level.

The Bray–Curtis dissimilarity measure was used to generate a microbial community distance matrix based on the most common bacterial phyla (relative abundance of Proteobacteria, Acidobacteria, Actinobacteria, Verrucomicrobia, Bacteroidetes, Chloroflexi, Cyanobacteria, Firmicutes, Gemmatimonadetes, Planctomycetes, and others). Similarly, the Euclidean distance was used to create a geographical distance matrix from latitude and longitude (in km). We considered that one degree is equivalent to 111.32 km for these analyses (Proffitt et al. 1989, Novembre et al. 2005, Chan and Wu 2015). With these matrices, the relationship between microbial turnover (i.e., changes in community composition from one community to another) and geographical distance (Lear et al. 2014) was evaluated. The significance of this relationship was assessed using Mantel test correlations. Microbial composition data were log-ratio-transformed before these analyses to reduce interdependence.

Primer sets.—We explored the effect of primer set choice on bacterial diversity and composition following three different approaches. First, independent analyses were conducted using soil samples collected from nine different locations: six ecosystem types located in the surroundings of Richmond (New South Wales, Australia; a bog, a natural forest, a grove, a grassland, a forest plantation, and a park) and three crops from Narrabri (New South Wales, Australia), Cunderdin (Western Australia, Australia), and Karoonda (South Australia, Australia). For each location, three soil samples were collected from two different depths (0–10 and 0–20 cm), resulting in 54 samples in total. These samples were sequenced using three different primer sets: 27F/519R (regions V1–V3), 341F/805R (V3–V4), and 515F/806R (V4; Lane 1991, Herlemann et al. 2011) using the Miseq Illumina (2X300 paired end reads) platform. We used these regions because they were the most widely used in the studies included in the database (58% of the soil samples analyzed). Initially, low-quality regions \((Q<20)\) were trimmed from the 3' end of the sequences using SEQTK (available online)\(^\text{10}\) and the paired ends were joined using FLASH (Magoc and Salzberg 2011). Primers were removed from the resulting sequences using SEQTK and a further round of quality control was subsequently conducted in MOTHUR (Schloss et al. 2009) to discard sequences with ambiguous characters or more than 8 homopolymers. Short sequences \(<400\text{ bp for primer sets }27\text{F}/519\text{R and }341\text{F}/805\text{R and }<250\text{ bp for primer set }515\text{F}/806\text{R}\) were also discarded. Operational Taxonomic Units (OTUs) were built at 97% sequence similarity using UPARSE (Edgar et al. 2013). Singleton sequences were discarded, as well as chimeric sequences identified by the UCHIME algorithm using the SILVA gold 16S rRNA gene reference database (Edgar et al. 2011).

Statistical analyses III: Meta-analysis cross-validation: In silico and experimental analyses

To ensure that our results were robust to potential sources of bias, a series of cross-validations using experimental data and in silico analyses were conducted to explicitly consider any bias on our estimates of bacterial diversity and composition resulting from the inclusion of studies using (1) various primer sets, (2) different top soil depths, (3) Shannon diversity indexes calculated at different levels of rarefaction, (4) different sequencing technology, and (5) analyses conducted in different years. Please note that the analyses conducted in this section, described in the paragraphs below, were done only when data was available (e.g., soil depth), so the number of soil samples may slightly change among different analyses.

\(^{10}\) https://github.com/lh3/seqtk
Differences in bacterial diversity with soil depth.—

Soil depth.—Differences in bacterial diversity with soil depth may bias our results by affecting the direct comparison among different studies; extra precautions were taken to consider this issue. First, data were collected from studies focusing on the first layer of soil (average soil depth in our database is 10.32 ± 4.60 cm [mean ± SD]). Second, additional analyses were performed to correct any bias arising from using different soil depths in our database. For doing so, partial correlations were conducted between latitude and altitude with Shannon diversity correcting for both the year of publication and the year of soil collection. In addition, to account for slight differences in the calculation for obtaining similarity cut-offs for the generation of OTUs (e.g., clustering method) have occurred during the last decade, the core of our results were repeated using partial correlations between latitude/longitude and Shannon diversity correcting for both the year of publication and the year of soil collection. In addition, and to account for slight differences in methodological approaches with time, the PERMANOVA analyses were repeated for bacterial composition, with geographical region as a fixed factor and the year of publication and that of soil collection as covariates.

Including primer sets and soil depth in statistical analyses.—The linear regression, Random Forest and SEM analyses explained in the Statistical analyses I section above were repeated using the residuals from a one-way ANOVA with primer set (8F-541R, 515F-816R, 338F-907R, others, and non-provided) as a fixed factor,
soil depth as a covariate, and bacterial diversity as a response variable (i.e., residuals of bacterial diversity). Moreover, the correlation analyses between spatial variables and the main bacterial taxa were repeated using residuals from one-way ANOVAs (see Statistical analyses II) with primer set as a fixed factor, soil depth as a covariate and main bacterial taxa as a response variable.

Finally, a variance partitioning linked to technical (primer sets and soil depth), environmental (soil C, C:N, pH, climate) and spatial (latitude and altitude) variables on bacterial diversity and composition was calculated using the package vegan from R package (Oksanen et al. 2015, see Tedersoo et al. 2014 for a similar approach). The main goal of this analysis was to quantify the relative importance of technical vs. environmental variables as predictors of bacterial diversity and composition in our database.

**RESULTS**

**Geographical patterns in diversity and composition**

Bacterial diversity, as measured with Shannon diversity, increased from Antarctica to the tropics, but remained similar across the Northern Hemisphere (Fig. 2a, b). In addition, we found a negative relationship between altitude and Shannon diversity (Fig. 2c). Moreover, similar spatial patterns in bacterial diversity were observed at both the soil sample and site levels (Appendix S1: Fig. S2). Shannon diversity was lower in the Southern than Northern Hemisphere for drylands, tundra, and temperate and continental grasslands, which accounted for ~38% of the sites included in the database (Fig. 3; P < 0.05). Differences in Shannon diversity between hemispheres were especially noticeable for drylands and temperate and continental grasslands (P < 0.01). However, we did not find any differences between the hemispheres for either agro-systems, or tropical and temperate forests (Fig. 3; P > 0.05).

As a whole, the tropical regions had the highest taxonomic similarity across sites (Fig. 4a). Bacterial composition also showed distinct patterns in both hemispheres, being dominated by Acidobacteria and Actinobacteria in the Northern and Southern Hemispheres, respectively (Fig. 4b–f; Appendix S1: Fig. S3). While Proteobacteria co-dominated both hemispheres, a sharp decrease in the relative abundance of this phylum was observed toward the Antarctic Circle (Fig. 4f). Similar patterns were observed at both the sample and site levels (Appendix S1: Fig. S3). As found with bacterial diversity, these patterns in bacterial composition (i.e., Actinobacteria, Acidobacteria, and Proteobacteria) were maintained for drylands, temperate forests, tundra and temperate and continental grasslands, and also for agro-systems (Appendix S1: Fig. S4; P < 0.05).
After ensuring that soil sample and site levels provided similar patterns, we felt it was reasonable to work at a sample-level scale for further modelling. Our Random Forest analyses explained 47% of the variance for bacterial diversity and indicated that climatic features such as precipitation seasonality, diurnal temperature range, aridity, and mean annual temperature were the most important predictors of bacterial diversity at a global scale (Fig. 5). Latitude, climate, soil pH, soil total C, and C:N ratio also showed a significant influence on the distribution of bacterial diversity (Fig. 5).

Our SEM model explained ~21% of the variance for bacterial diversity and indicated that climatic features such as precipitation seasonality, diurnal temperature range, aridity, and mean annual temperature were the most important predictors of bacterial diversity at a global scale (Fig. 5). Latitude, altitude, soil pH, soil total C, and C:N ratio also showed a significant influence on the distribution of bacterial diversity (Fig. 5).

Microbial turnover along geographical distances

Our results indicated that bacterial community turnover may follow different patterns in the Northern and Southern Hemispheres. Thus, while bacterial community dissimilarity increased with geographical distance in the Southern Hemisphere (Fig. 9b), suggesting turnover limitation, this pattern was much weaker, nonetheless significant, in the Northern Hemisphere (Fig. 9a).

Correlations between bacterial composition and environmental drivers

Acidobacteria and Proteobacteria were positively related to latitude, while the opposite pattern was observed for Chloroflexi and Actinobacteria (Table 1; Appendix S1: Table S1; P < 0.001). Interestingly, distance from equator (absolute latitude) was much less related to this distribution pattern than latitude per se (Table 1; Appendix S1: Table S1). Similarly, soil total C and C:N ratio were positively related to Acidobacteria and Proteobacteria, but negatively related to Chloroflexi and Actinobacteria (Table 1; P < 0.05). Moreover, Acidobacteria and Proteobacteria were negatively related to soil pH, which was positively related to Chloroflexi and Actinobacteria (Table 1; P < 0.05). Aridity had a positive effect on Firmicutes, Chloroflexi, and Actinobacteria, but a negative effect on Proteobacteria (Table 1; P < 0.05). Finally, diurnal temperature range had a positive effect on Firmicutes, Chloroflexi, and Actinobacteria, but was negatively related to Proteobacteria (Table 1; P < 0.05).

Meta-analysis cross-validation

Primer sets.—Using the soils collected from different ecosystems in Australia, we did not find any significant differences among primer sets for bacterial Shannon diversity at the OTU level (P > 0.05; Appendix S1: Fig. S7). However, we found minor but significant differences between primer sets for bacterial composition (P < 0.05;
These differences were mainly occurring between regions V1–V4 (27F/519R and 341F/805R) and V4 alone (515F/806R). Additional analyses indicated that primer sets 27F/519R and 341F/805R provided similar relative abundance for main bacterial phyla (but not for Proteobacteria; Appendix S1: Table S2). Consequently, the results related to Proteobacteria need to be interpreted with caution.

Regarding in silico analyses, we did not find any significant differences among primer sets for bacterial Shannon diversity at the OTU level or composition ($P > 0.05$), regardless of the RDP subset considered (Figs. S9 and S10 in Appendix S1). Together with the sequencing analyses, these results provide evidence that using different primer sets do not bias the results on the relative abundance of phyla and bacterial diversity, and thus the results reported are robust to the different primer sets used by the studies summarized in our database. Despite these analyses, and because our primer selection only involved regions V1–V3, V3–V4, and V4 of rRNA 16S rDNA, we repeated our core analyses, using regions V1–V3, V3–V4, and V4 for bacterial diversity and V1–V3 and V3–V4 for bacterial composition; no significant differences were found between primer sets for Shannon diversity or composition (as supported by the analyses of soil samples from Australia and in silico results). We also repeated these analyses using the most common primer sets for bacterial composition found in this meta-analysis (Fig. S11 in Appendix S1). Our results were maintained after accounting for these issues, as we still found negative statistically significant relationships between altitude and bacterial diversity, and a reduction in bacterial diversity from the tropics to Antarctica ($P < 0.01$; Appendix S1: Fig. S12). Similarly, we found a dominance of Acidobacteria and Actinobacteria in the Northern and Southern hemispheres, respectively (Appendix S1: Fig. S12).
Soil depth.—We did not find any significant differences between soil depth (0–10 and 0–20 cm) in our experimental data, either for bacterial diversity or composition at the phyla level and regardless of the primer set used ($P > 0.05$; Appendix S1: Figs. S7 and S8). These results were maintained across different ecosystem types (Appendix S1: Figs. S7b and S13). In addition, partial correlation analyses between latitude/altitude and bacterial diversity accounting for soil depth provided further evidence that these results were not biased by using different soil depths in our database (latitude, Shannon index, Pearson’s $r = 0.187$, $P = 0.001$; altitude, log$_{10}$-transformed).

Shannon index, $r = -0.159$, $P = 0.002$).

**Shannon diversity calculated at different rarefaction levels.**—Increased sequencing depth (i.e., number of reads in the different studies) biased our data set by artificially increasing the values of the Shannon index (Pearson’s $r = 0.209$, $P < 0.001$; $n = 349$ soil samples). However, partial correlations between latitude/altitude and bacterial diversity correcting for number of sequences used to calculate this index (after rarefaction) showed similar results to those presented in Fig. 1: positive/negative relationship between latitude/altitude and bacterial diversity (latitude, Shannon index, $r = 0.259$, $P < 0.001$; altitude, log$_{10}$-transformed Shannon index, $r = -0.169$, $P = 0.002$, $n = 346$). In addition, after using regression tree analyses, we found a threshold at 8913 sequences. Shannon diversity calculated below 8913 sequences/sample still had a significant relationship with the number of sequences used for its calculation (Pearson’s $r = 0.181$, $P = 0.003$; $n = 277$ soil samples). However, this was not the case for Shannon diversity calculated above

---

**Fig. 5.** Importance of soil bacterial diversity drivers, as measured with the percentage of increase of mean square error (MSE) in the Random Forest analysis. Results are shown for the whole data set ($n = 390$ study cases). Significance levels of each predictor are *$P < 0.05$, **$P < 0.01$.

**Fig. 6.** Structural equation model describing the effects of multiple drivers on bacterial diversity. (a) Numbers adjacent to arrows are indicative of the effect size of the relationships. Continuous and dashed arrows indicate positive and negative relationships, respectively. $R^2$ denotes the proportion of variance explained. For graphical simplicity, letters on arrows mean (A) Spatial $\rightarrow$ Aridity $= 0.14$, Spatial $\rightarrow$ Mean annual temperature $= -0.90$$^*$*, Spatial $\rightarrow$ Diurnal temperature range $= -0.13$$^*$*, (B) Altitude $\rightarrow$ Aridity $= -0.01$, Altitude $\rightarrow$ Mean annual temperature $= -0.20$$^*$*, Altitude $\rightarrow$ Diurnal temperature range $= 0.38$$^*$*, Altitude $\rightarrow$ precipitation seasonality $= 0.13^*$. Hexagons are composite variables. Squares are observable variables. Panel (b) represents standardized total effects (direct plus indirect effects) derived from the structural equation model used. The numbers in the parentheses are bootstrap P-values.
this threshold. Thus, Shannon diversity and the number of sequences used for its calculation were no longer strongly related ($r = -0.159; P = 0.182; n = 274$). Most importantly, the observed latitudinal and altitudinal patterns in this study followed similar trends even when controlling for potential biases caused by differences in sequencing effort among studies (latitude, Shannon index, $r = 0.414, P < 0.001$; altitude, log$_{10}$-transformed Shannon index, $r = -0.191, P = 0.108; n = 72$), especially considering the low number of samples included in these analyses ($n = 72$).

Technology and year of publication/collection.—Both 454 (97% of data) and Miseq Illumina (3% of data) showed similar levels of bacterial diversity in this study ($P > 0.05$; Appendix S1: Fig. S14). Finally, the relationship between latitude/altitude and bacterial diversity was still maintained after correcting for the year of publication and soil collection (latitude, Shannon index, $r = 0.263, P < 0.001$; altitude, log$_{10}$-transformed Shannon index, $r = -0.138, P = 0.025$). In addition, we found significant differences among geographical regions for bacterial diversity and composition after including the year of publication and soil collection in our PERMANOVA analyses as covariables ($P < 0.001$).

Additional statistical analyses.—Our main finding, a reduction and increase in bacterial diversity with altitude and latitude, respectively, was maintained after using the residuals derived from a one-way ANOVA with primer set (8F-541R, 515F-816R, 338F-907R, others, and non-provided) as a fixed factor and soil depth as a covariate (Appendix S1: Fig. S2). This applies to analyses conducted at both sample and site levels (Appendix S1: Fig. S2). In addition, we found similar results after repeating our Random Forest and SEM analyses (Figs. 5 and 6 vs. Appendix S1: Figs. S15 and S16) using the residuals of bacterial diversity as a response variable. Our Random Forest still explained ~40% of the variance found in this variable, and latitude, altitude, and climatic features were still the most important factors controlling its variation, followed by soil properties (Fig. 5 vs. Appendix S1: Fig. S15). SEM analyses conducted using these residuals also showed similar patterns to those obtained when using bacterial diversity as a dependent variable (Fig. 6 vs. Appendix S1: Fig. S16).

Finally, our variance partitioning analyses highlighted that both spatial and environmental variables together explained a higher portion of the variance than technical variables (primer sets and soil depth) for both bacterial diversity and composition (Appendix S1: Fig. S17). Similar results were found at the sample and site levels (Appendix S1: Fig. S17).

**DISCUSSION**

Our study provides evidence of similarities and discrepancies in the spatial patterns of micro- and macroorganisms in terrestrial ecosystems. For instance, soil bacterial diversity increased from Antarctica to the equator (MacArthur 1975, Rohde 1992, Begon et al. 2006); however, this diversity did not differ systematically with latitude in the Northern Hemisphere, and even tended to increase in the Arctic region. Soil bacterial composition also showed distinctive patterns among hemispheres, being dominated by Acidobacteria and Actinobacteria in the Northern and Southern Hemispheres, respectively. However, Proteobacteria was co-dominant in both hemispheres. Moreover, we found a reduction in soil bacterial diversity with altitude. Remarkably, our SEM and Spearman correlation analyses provided strong evidence that distance from equator (absolute latitude) had a much lower influence on bacterial diversity and composition than latitude per se. While climate correlated with diversity as it varied with altitude (i.e., high diurnal temperature range and low temperature correspond with low diversity at high altitudes), geographical gradients in soil total C and species turnover were important drivers of the observed latitudinal patterns. These distinct latitudinal and altitudinal patterns included substantial variation, but were maintained even when controlling for multiple potential biases related to
different primer sets and 16 rRNA regions, soil depths, rarefaction efforts, sequencing techniques and the year of publication used by the studies included in our database (see Results: Meta-analysis cross-validation).

Previous studies have found either no relationship (Fierer et al. 2011) or a negative relationship (Lyngwi et al. 2013) between altitude and soil bacterial diversity. Factors such as the high dispersal capacity of bacterial communities (Fierer et al. 2011) may have obscured this spatial pattern in shorter altitudinal gradients and in local-scale studies. However, we provide evidence that, as previously observed with plants and animals from

Fig. 8. Selected pairwise relationships between latitude, soil total C (soil C, measured as the percentage of Soil mass), and soil bacterial diversity. Data in panel (a) represent means ± SE; Antarctic (n = 17), middle latitude south (n = 43), equator (n = 44), middle latitude north (n = 267), and Arctic (n = 19). The different letters in panel (a) indicate significant differences among latitudinal ranges and solid lines in panels (b), (c), (d), and (e) represent the fitted linear regressions.
mid to high elevations (MacArthur 1975, Rohde 1992, Rahbek 2005, Begon et al. 2006, Sundqvist et al. 2013), a negative relationship between altitude and bacterial diversity can be found at the global scale when considering a wide altitudinal range (0–4600 m). Diurnal temperature range and mean annual temperature were significant factors related to bacterial diversity in our study. Our results indicate that the observed negative effect of altitude on soil bacterial diversity can be, at least in part, indirectly attributed to the high diurnal temperature range and low temperatures found in high-elevation areas. The increased physiological stress associated with rapidly changing environmental conditions, and the decrease in net primary productivity near to the mountain tops, can constrain bacterial diversity, as previously observed with plants and animals (MacArthur 1975, Rohde 1992, Begon et al. 2006, Young et al. 2013). An important finding of our study is that climate itself cannot fully account for the latitudinal patterns observed in soil bacterial diversity. The equator divides the global climate and soil pH patterns roughly into two mirror images. However, this did not happen for either bacterial diversity or soil total C, which increased from the Antartic to the Northern Hemisphere. Interestingly, soil total C was positively related to soil bacterial diversity in the different analyses conducted, and was one of the most important drivers of the variation observed in this study. In agreement with our results, an increase of total and microbial biomass C from southern to northern latitudes has been previously reported (Xu et al. 2013, Wieder et al. 2014), and the importance of organic matter as a driver of soil bacterial diversity at regional and global scales has also been recently highlighted (Siciliano 2014, Maestre et al. 2015).

Our results provide strong evidence for a disconnection between the diversity patterns of higher organisms (plants and animals) and soil bacterial diversity in the Northern Hemisphere, which may be related to the differences in energy requirements and dispersal limitations of these organisms. Unexpectedly high levels of bacterial diversity have been previously reported at northern high latitudes (Steven et al. 2013), although the main drivers of this pattern remain unclear. Similarly, a recent global study focused on drylands found a positive effect of latitude on the diversity of soil bacteria (see Fig. 3 in Maestre et al. 2015). Contrary to plants and animals, many soil bacteria primarily rely on organic matter decomposition to obtain energy (Allison et al. 2010, Trivedi et al. 2013). Thus, the

### Table 1. Correlation coefficients (Spearman’s ρ) between latitude, distance from the equator, altitude, climate, and soil properties with the relative abundance of the main bacterial phyla (n = 588 samples).

<table>
<thead>
<tr>
<th>Factors</th>
<th>Acidobacteria</th>
<th>Actinobacteria</th>
<th>Chloroflexi</th>
<th>Firmicutes</th>
<th>Proteobacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latitude</td>
<td>0.162 (&lt;0.001)</td>
<td>−0.171 (&lt;0.001)</td>
<td>−0.105 (0.011)</td>
<td>−0.008 (0.852)</td>
<td>0.150 (&lt;0.001)</td>
</tr>
<tr>
<td>Distance from equator</td>
<td>−0.073 (0.078)</td>
<td>0.092 (0.023)</td>
<td>−0.083 (0.045)</td>
<td>−0.074 (0.072)</td>
<td>−0.012 (0.775)</td>
</tr>
<tr>
<td>Altitude</td>
<td>0.171 (&lt;0.001)</td>
<td>0.023 (0.578)</td>
<td>−0.001 (0.986)</td>
<td>−0.006 (0.884)</td>
<td>−0.003 (0.945)</td>
</tr>
<tr>
<td>Mean annual temperature</td>
<td>0.000 (0.997)</td>
<td>−0.018 (0.656)</td>
<td>−0.010 (0.817)</td>
<td>0.116 (0.001)</td>
<td>0.021 (0.606)</td>
</tr>
<tr>
<td>Diurnal temperature range</td>
<td>−0.013 (0.756)</td>
<td>0.150 (&lt;0.001)</td>
<td>0.097 (0.018)</td>
<td>0.204 (&lt;0.001)</td>
<td>−0.166 (&lt;0.001)</td>
</tr>
<tr>
<td>Precipitation seasonality</td>
<td>−0.067 (0.103)</td>
<td>0.073 (0.075)</td>
<td>0.284 (&lt;0.001)</td>
<td>−0.100 (0.016)</td>
<td>−0.141 (&lt;0.001)</td>
</tr>
<tr>
<td>Aridity</td>
<td>−0.192 (&lt;0.001)</td>
<td>0.291 (&lt;0.001)</td>
<td>0.242 (&lt;0.001)</td>
<td>0.162 (&lt;0.001)</td>
<td>−0.284 (&lt;0.001)</td>
</tr>
<tr>
<td>pH</td>
<td>−0.424 (&lt;0.001)</td>
<td>0.204 (&lt;0.001)</td>
<td>0.345 (&lt;0.001)</td>
<td>0.248 (&lt;0.001)</td>
<td>−0.269 (&lt;0.001)</td>
</tr>
<tr>
<td>Total C</td>
<td>0.219 (&lt;0.001)</td>
<td>−0.222 (&lt;0.001)</td>
<td>−0.097 (0.018)</td>
<td>−0.195 (&lt;0.001)</td>
<td>0.157 (&lt;0.001)</td>
</tr>
<tr>
<td>C : N</td>
<td>0.238 (&lt;0.001)</td>
<td>−0.160 (&lt;0.001)</td>
<td>−0.244 (&lt;0.001)</td>
<td>−0.150 (&lt;0.001)</td>
<td>0.179 (&lt;0.001)</td>
</tr>
</tbody>
</table>

*Note: P values are in brackets.*
lower soil total C content found in the Southern vs. the Northern Hemisphere may explain, at least in part, the observed diversity patterns. This relationship is especially noticeable in the Southern Hemisphere, which harbors some of the ecosystems with the lowest soil total C availability on Earth (Xu et al. 2013, Wieder et al. 2014), and suggests that soil total C may only limit the diversity of bacteria under very low C conditions. The overall positive relationship between soil total C and bacterial diversity found in this study may be part of the mechanisms underlying global ecosystem functioning and climate regulation (Wieder et al. 2014). For instance, the presence of large soil organic matter stocks and highly diverse bacterial communities in northern latitude regions (Xu et al. 2013, Wieder et al. 2014) suggest that increases in temperature may be followed by enhanced organic matter decomposition and increased soil respiration, with large implications for C fluxes and climate regulation at the global scale (Xu et al. 2013, Wieder et al. 2014). Interestingly, the highest differences in bacterial diversity between the hemispheres were found in C-poor and low productivity biomes such as those from the Antarctic Circle, drylands, and grasslands. These results suggest that, in the nutrient-depleted soils from the Southern Hemisphere, reductions in productivity and associated C inputs to the soil linked to climatic events such as droughts, which are likely to be more recurrent under ongoing climate change (Dai 2012, IPCC 2013), can have a significant negative impact on soil bacterial diversity. Other than soil total C, lack of vascular plants and extreme climatic conditions can also explain the lowest bacterial diversity found in the Antarctic compared to Arctic regions. For instance, the climate at similar latitudes in the Antarctic region is much more extreme than in the Arctic, with a lower annual mean temperature in the former (~11.0°C vs. ~4.5°C in our data set; P = 0.022). This may further limit bacterial diversity in the Antarctic sites.

Interestingly, while we did not find any direct effect of altitude on bacterial diversity in our SEM, we still found a strong and significant positive direct effect of latitude on bacterial diversity. These results suggest that while climate and soil properties suffice to explain all the effects from altitude on bacterial diversity, we were unable to determine all the potential mechanisms explaining the positive effect of latitude on bacterial diversity. As a complementary explanation to the lowest bacterial diversity found in the Southern Hemisphere, the increase in bacterial community dissimilarity with geographical distance found in the Southern Hemisphere suggests that bacterial turnover may be lower in this hemisphere (vs. Northern Hemisphere), which may reduce the probability of phyla dispersion and colonization. Contrary to this, we found a weak, nonetheless significant, relationship between community dissimilarity and geographical distances in the Northern Hemisphere. The fact that this hemisphere has an enormous and almost continuous mass of land, while the continents in the Southern Hemisphere (e.g., Australia, Antarctica, South America, and Africa) are separated by large water bodies and have less land mass, may also contribute to explain the lowest bacterial diversity found in the latter. Remarkably, these results challenge the extended idea that microbes are not limited by dispersion mechanisms, and suggest that the island biogeography theory may also potentially apply to microorganisms at the global scale (MacArthur 1975, Rohde 1992, Begon et al. 2006). While this idea has been elaborated for at the species level (Barberán et al. 2014), our results indicate that this may also apply to phylum-level composition of bacteria. Interestingly, and contrary to what has been found with plants and animals, the tropical and Antarctic regions had the highest and lowest taxonomic similarity across sites, respectively. This pattern is perhaps linked to the warm/cold environmental conditions and high/low spatial connectivity across sites characterizing the tropical and Antarctic regions, respectively (MacArthur 1975, Rohde 1992, Begon et al. 2006).

It is also important to note that the dominant bacterial communities differed in the Northern and Southern Hemispheres (Figs. 4 and Appendix S1: Fig. S2). Ancient phyla such as Actinobacteria and Chloroflexi (Battistuzzi and Hedges 2009) were more abundant in the Southern Hemisphere (Table 1), where some of the oldest, nutrient-depleted soils on Earth are found (Johnson 2009). These bacterial phyla have been shown to possess important adaptations that enable them to resist environmental harshness (e.g., ability to survive desiccation and low nutrient availability conditions; Battistuzzi and Hedges 2009, Trivedi et al. 2013). In addition, Actinobacteria have been suggested to effectively scavenge nutrients in nutrient-poor ecosystems, giving this group an advantage in these environments (Zechmeister-Boltenstern et al. 2015). In contrast, the Northern Hemisphere was dominated by relatively old bacterial phyla such as Acidobacteria and Proteobacteria (Battistuzzi and Hedges 2009), which are highly diverse groups of microorganisms linked to organic C degradation (Fierer et al. 2007, Trivedi et al. 2013). These phyla may have evolved in relatively old soils with high nutrient availability, after the accumulation of C and N via atmospheric fixation from other ancient groups such as Cyanobacteria and Chloroflexi (Battistuzzi and Hedges 2009). A decrease in soil bacterial diversity and new phyla such as Acidobacteria and Proteobacteria (Battistuzzi and Hedges 2009, Trivedi et al. 2013), together with an increase in ancient bacterial phyla such as Actinobacteria, has been previously observed with soil development in old soils (Jangid et al. 2014). However, the mechanisms that control these processes are uncertain. Gradients in the availability of nutrients such as nitrogen and phosphorus from northern latitudes to the Southern Hemisphere linked to the last glaciation (Johnson 2009) may also explain the observed global patterns in bacterial diversity and composition. However, we have not been able to explicitly assess the relationship with these nutrients because of a lack of data in the primary studies compiled. Supporting this idea, the highest differences in the relative abundance of Actinobacteria, Acidobacteria, and Proteobacteria between
the hemispheres are found in nutrient-poor environments such as drylands and the tundra (Pointing et al. 2009).

Our results support previous studies highlighting the importance of factors such as pH (Fierer and Jackson 2006, Fierer et al. 2009, Lauber et al. 2009), spatial influence (Horner-Devine et al. 2004, Martiny et al. 2011, Hanson et al. 2012, Maestre et al. 2015) and soil total C (Siciliano 2014, Maestre et al. 2015) as drivers of changes in soil bacterial diversity and composition. We also acknowledge that the explanatory ability of our statistical models was relatively low, something that is not surprising given the nature of the data used. Altogether, our study, which encompasses a wide range of ecosystem types, climates, and soil properties, provides a reasonable initial snapshot of the patterns and mechanisms driving bacterial diversity at the global scale. However, we acknowledge the limitations of our analytical approach. Meta-analyses are always noisier than field studies using standardized survey and analytical protocols because of the methodological differences among the studies synthesized. As such, meta-analyses in ecology have been encouraged because of the general belief that differences in soil depth, primer sets and sequencing depth can potentially alter the values obtained for both microbial composition and diversity (Engelbrektsen et al. 2010, Cai et al. 2013, Fredriksen et al. 2013). However, we argue that next generation sequencing data are noisier than any other soil data that have been synthesized using meta-analysis, such as microbial stoichiometry (Xu et al. 2013), litter decomposition rates (García-Palacios et al. 2015), microbial abundance (Fierer et al. 2009, Serna-Chávez et al. 2013), fungal:bacterial ratios (Fierer et al. 2009) and soil bacterial diversity estimated using terminal restriction fragment length polymorphism (T-RFLP; Lozupone and Knight 2007). Here, we provided solid evidence from both in silico and experimental approaches that technical aspects such as the primer set used or the depth at which samples were collected did not alter the main global patterns in bacterial communities reported here. Moreover, we found that environmental and spatial variables explained levels of the variance in soil bacterial diversity and composition similar to those reported by Tedersoo et al. (2014), who studied soil fungal communities using data gathered from a standardized global survey. As such, our study provides a novel conceptual framework integrating multiple hypotheses to build a system-level understanding of the global drivers of bacterial diversity in terrestrial ecosystems. More importantly, our combined approach of ecological theory and novel statistical modeling could be highly useful for future studies addressing global patterns in microbial diversity in terrestrial ecosystems by providing multiple tools to identify and address existing limitations related to our understanding of global patterns in soil bacterial diversity and composition. Further, our findings provide a testable hypothesis and thus open the door to conducting a large scale global study to test whether the global patterns we found through synthesized data remain the same with original data collected with a standardized method. As such, our work provides the first step toward understanding global patterns and drivers of bacterial diversity and composition.

In summary, we found that soil bacterial diversity and composition follow different patterns related to latitude and temperature in Northern vs. Southern Hemispheres, in distinct contrast to what has been found with plant and animal diversity. Moreover, our results indicate that bacterial diversity decreases with altitude in terrestrial ecosystems. While extreme climatic conditions were the main drivers of change in soil bacterial diversity with altitude, factors such as the low soil C content and microbial turnover rates best explained the low soil bacterial diversity found in the Southern Hemisphere. Thus, our findings suggest a paradigm shift in our understanding of the biogeographical patterns and drivers of soil bacterial diversity by identifying contrasting biodiversity patterns to those previously observed with plants and animals; and the possible mechanisms that control it at a global scale.

Acknowledgments

We acknowledge the use of data from NEON, which are made available subject to the NEON Data Policy. We also thank Cateriona Macdonald and Loïc Nazaries for assistance in data gathering, Melissa S. Martin and Jasmine Grinyer for revising the English of this manuscript. This research is supported by the ARC project DP13010484. F. T. Maestre is supported by the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement 242658 (BIOCOM) and by the Spanish Ministry of Economy and Competitiveness (BIOMOD project, CGL2013-44661-R). The authors declare no competing financial interests.

Literature Cited


He, Y., et al. 2013. Comparison of microbial diversity determined with the same variable tag sequence extracted from two different PCR amplicons. BMC Microbiology 16:208.


IPCC. 2013: Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK and New York, New York, USA.


**Supporting Information**

Supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecm.1216/suppinfo

**Data Availability**

Data associated with this paper has been deposited in figshare: https://figshare.com/s/08c8117230171e5803f06ec4bbcf141