# Palaeoclimate explains a unique proportion of the global variation in soil bacterial communities

**Article** · August 2017

DOI: 10.1038/s41559-017-0259-7



**10 authors**, including:

Manuel Delgado-Baquerizo
University of Colorado Boulder
**104** PUBLICATIONS **1,597** CITATIONS
SEE PROFILE

Andrew Bissett
The Commonwealth Scientific and Industrial …
**67** PUBLICATIONS **1,435** CITATIONS
SEE PROFILE

David J Eldridge
UNSW Sydney
**220** PUBLICATIONS **6,301** CITATIONS
SEE PROFILE

Fernando T Maestre
King Juan Carlos University
**291** PUBLICATIONS **12,569** CITATIONS
SEE PROFILE

nature ecology & evolution

ARTICLES

DOI: 10.1038/s41559-017-0259-7

# Palaeoclimate explains a unique proportion of the global variation in soil bacterial communities

Manuel Delgado-Baquerizo[1,2*], Andrew Bissett[3], David J. Eldridge[4], Fernando T. Maestre[5], Ji-Zheng He[6,7], Jun-Tao Wang[6], Kelly Hamonts[2], Yu-Rong Liu[6], Brajesh K. Singh[2,8] and Noah Fierer[1,9]

**The legacy impacts of past climates on the current distribution of soil microbial communities are largely unknown. Here, we use data from more than 1,000 sites from five separate global and regional datasets to identify the importance of palaeoclimatic conditions (Last Glacial Maximum and mid-Holocene) in shaping the current structure of soil bacterial communities in natural and agricultural soils. We show that palaeoclimate explains more of the variation in the richness and composition of bacterial communities than current climate. Moreover, palaeoclimate accounts for a unique fraction of this variation that cannot be predicted from geographical location, current climate, soil properties or plant diversity. Climatic legacies (temperature and precipitation anomalies from the present to ~20 kyr ago) probably shape soil bacterial communities both directly and indirectly through shifts in soil properties and plant communities. The ability to predict the distribution of soil bacteria from either palaeoclimate or current climate declines greatly in agricultural soils, highlighting the fact that anthropogenic activities have a strong influence on soil bacterial diversity. We illustrate how climatic legacies can help to explain the current distribution of soil bacteria in natural ecosystems and advocate that climatic legacies should be considered when predicting microbial responses to climate change.**

The climate of a particular region varies over time, often resulting in large-scale biome migrations that drive the current distribution of plant communities[1–4]. For example, long-term climatic legacies have shaped the distribution and diversity of plant communities in terrestrial ecosystems through dispersal-limited recolonization and environmental filtering[2–4]. Similarly, long-term regional climate history could conceivably explain significant proportions of the variation found in the current richness and composition of soil microbial communities. For example, a recent study provides indirect evidence that the last glaciation may have influenced the current distribution of strains of the soil bacterial genus *Streptomyces* across the United States[5,6]. However, the broader role of past climate conditions in regulating the current distribution of microbial communities remains largely unexplored[5]. If past climates help to explain the current distribution of microbial communities, careful consideration of climatic legacies could improve our capacity to predict how soil microbial communities will respond to forecasted climate changes, and how this response will affect the myriad ecosystem services that they provide (such as decomposition, nutrient cycling and climate regulation)[7–9].

In theory, palaeoclimate could explain the current distribution of soil microbial communities directly, through differential changes in temperature and precipitation patterns across millennia[5,6]. Soil bacteria are known to have short generation times leading to a fast turnover rate, but they are also highly sensitive to changes in temperature[10]. For example, a recent study demonstrated that a wide range of soil bacterial taxa show predictable and consistent preferences for particular temperature conditions[10]. These intrinsic characteristics of microbial communities surely influence their direct response to palaeoclimate. For instance, the community composition of fast-growing invertebrates responded immediately to large and abrupt changes in temperature after the most recent glaciation, a response that left a strong signature in their contemporary distribution[11]. Likewise, abrupt changes in climate, which may have occurred before 10,000 years ago[12], might have also left a strong signature on the structure of soil bacterial communities. In this respect, a direct effect from palaeoclimate on soil microbial communities might have occurred in the past (for example in response to a severe climatic event), but the consequences of this rapid compositional shift might still be detectable today.

Palaeoclimate can also influence the current structure of bacterial communities indirectly, through its influence on soil properties and plant community structure[13–15]. Thus, variations in soil properties such as pH and total organic carbon, which can have strong effects on microbial distributions[13–15] and change slowly during ecosystem development[16,17], could drive the effects of palaeoclimate on the contemporary patterns of microbial community composition and richness. Likewise, palaeoclimate effects on plant communities[2–4] may be associated with corresponding changes in the composition of soil microbial communities[18]. Although the growing literature focuses on the main drivers of soil microbial communities in terrestrial ecosystems, we do not know whether climatic legacies contribute to their current richness and composition patterns at regional or global scales.

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA. [2]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales 2751, Australia. [3]CSIRO, Oceans and Atmosphere, Hobart, Tasmania 7000, Australia. [4]Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia. [5]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. [6]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China. [7]Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia. [8]Global Centre for Land Based Innovation, Western Sydney University, Building L9, Locked Bag 1797, Penrith South, New South Wales 2751, Australia. [9]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA. *e-mail: M.DelgadoBaquerizo@gmail.com

Additionally, if climatic legacies play an important role in regulating current soil microbial distribution, agricultural practices may reduce or remove any potential effects of palaeoclimate on microbial community composition and richness. Agricultural practices are known to alter soil microbial communities directly, for instance by introducing new bacterial taxa associated with crop rhizospheres[19] or through fertilization[20], and indirectly, through changes in soil properties (soil carbon, soil pH and microbial communities)[21]. Marked changes in the composition and richness of soil bacteria derived from agricultural practices might potentially mitigate the direct and indirect influences of climatic legacies on soil microbial communities via soil properties. Soil disturbance is expected to increase exponentially this century, owing to the increasing intensification of agricultural production needed to meet an increase in demand for food by 70– 100% by 2050 (ref. [22]). Thus, understanding how agricultural intensification will shift the signature of climatic legacies on microbial communities could improve our ability to evaluate and manage anthropogenic soil disturbances.

Here, we evaluated the relative importance of palaeoclimate and current climate as predictors of the richness (number of phylotypes observed per sample) and composition (relative abundance of phylotypes) of soil bacteria at global and regional scales after accounting for key drivers of bacterial distribution such as geographical location, soil properties and plant diversity. We did so using data from five separate regional and global datasets including information on the structure of bacterial communities assessed by 16S ribosomal RNA gene sequencing (see Methods). Together, these datasets included more than 1,000 sites from all continents except Antarctica, covering a broad range of ecosystem types (see Methods and Supplementary Fig. 1). We tested the following hypotheses: (i) palaeoclimate predicts a unique portion of the variation in the current richness and composition of soil bacterial communities in terrestrial ecosystems; (ii) climatic legacies (measured as the temperature and precipitation anomalies[12] between an estimate of climate 20,000 years ago and another estimate for the present day) affect the structure of current bacterial communities both directly and indirectly through soil properties and plant diversity; and (iii) soil disturbances linked to agricultural production reduce the relative importance of palaeoclimate as a predictor of current microbial community richness and composition. It was not our intention to merge the five datasets used, which vary in sampling design and experimental methods (such as primer sets), but to test our hypotheses using five independent regional and global datasets from ecosystems that differed in their vegetation, climate and soil attributes (Methods; Supplementary Fig. 1).

## Results

We first used variation partitioning[23] to quantify the relative contribution of past and current climates as predictors of the richness and composition of soil bacterial communities. We also included soil properties and spatial variables[15] in our models. This approach allowed us to quantify the unique contribution of climate from a particular period to explain the current distribution of soil bacteria and to differentiate this contribution from that shared among all predictors. Environmental drivers such as plant diversity, soil properties and geographical location explained unique portions of the variation in soil bacterial richness and composition in all datasets (Fig. 1, Supplementary Figs. 2 and 3). Most importantly, climate variables from mid-Holocene and Last Glacial Maximum climates explained a unique percentage of the variation in the richness and composition of soil bacterial communities (Fig. 1, Supplementary Figs. 2 and 3). Overall, palaeoclimate was a better predictor of soil bacterial richness and composition than current climate in all five datasets (Fig. 1), suggesting that models using current climate alone have a limited predictive power at regional to global scales. Palaeoclimate also shared a large part of the variance explaining bacterial community richness and composition with plant richness and/or soil properties, suggesting that a large fraction of the apparent effects of palaeoclimate on soil bacterial communities may be driven by its direct and indirect effects on these ecosystem variables.

We then used structural equation modelling (SEM; see Methods) to assess the role of climatic legacies in driving bacterial community composition and richness, and to separate direct effects (temperature and precipitation anomalies between an estimate of climate

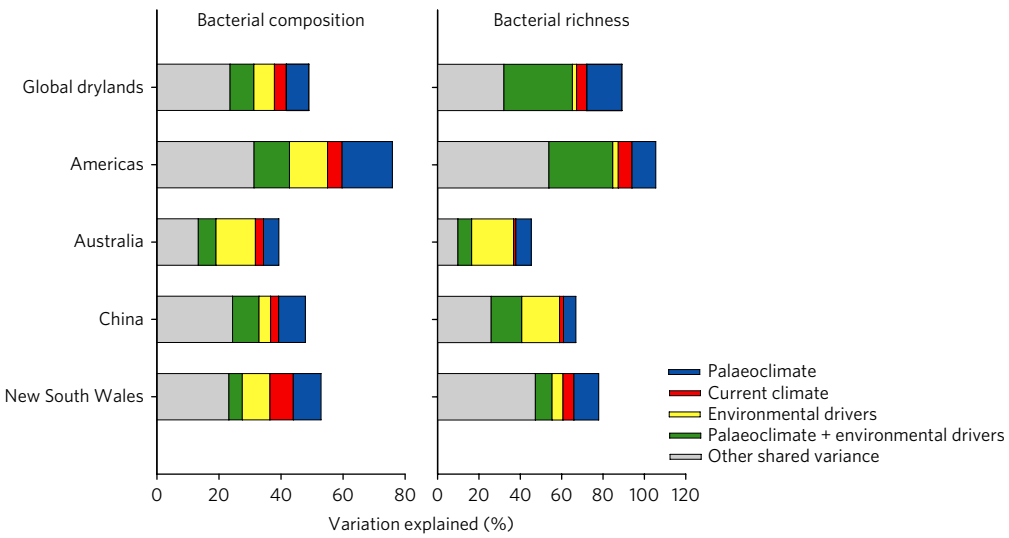


**Fig. 1 | Relative contribution of the different predictors used to model bacterial composition and diversity.** Panels represent results from variation partitioning modelling aiming to identify the percentage variance of bacterial community composition explained by past and current climate variables across five independent large-scale datasets. Unique and shared variance from the Last Glacial Maximum and mid-Holocene in predicting bacterial community composition and richness were merged in this figure for simplicity. Note that the variation explained by 'palaeoclimate + environmental drivers' is additional to the one explained either by palaeoclimate only or environmental only. An alternative version of this figure showing the unique and shared variance of each group of predictors can be found in Supplementary Figs. 2 and 3. Further information on the datasets used in these analyses can be found in refs [15,28,29,35,38].

20,000 years ago and another estimate for the present day) and indirect effects (via soil properties and plant diversity) of such legacies on soil microbial communities. Unlike regression analyses, SEM offers the ability to separate multiple pathways of influence and to investigate the complex relationships among environmental predictors commonly found in terrestrial ecosystems (Methods). As SEM

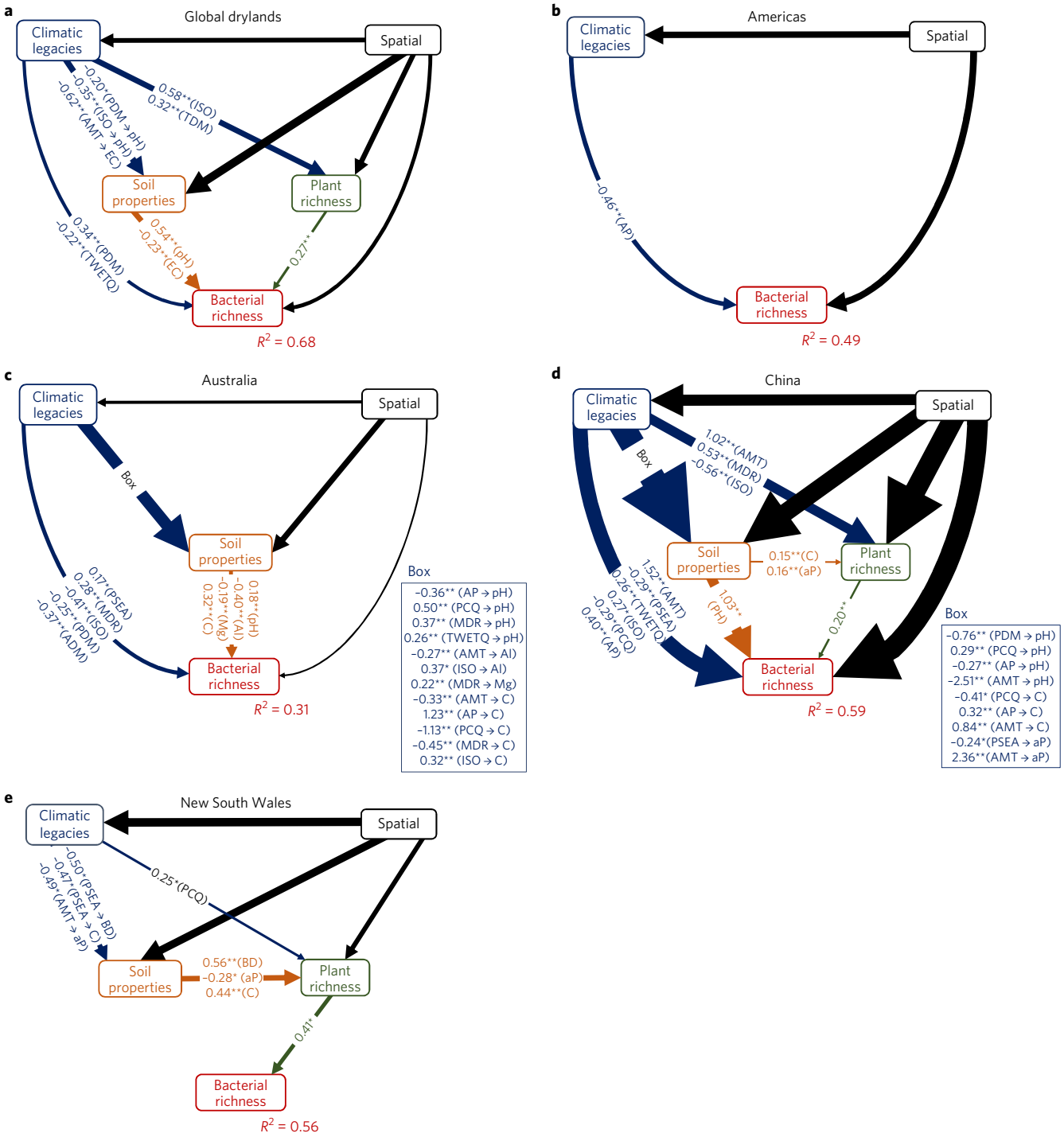


**Fig. 2 | Structural equation model accounting for the direct and indirect (plant diversity and/or soil properties) effects of climatic legacies on the diversity of bacteria across the five datasets used. a–e**, Numbers adjacent to arrows are path coefficients (*P* values) and are indicative of the standardized effect size of the relationship. Spatial influence (latitude and longitude) was included to control spatial autocorrelation; however, in this case, path coefficients were not included for simplicity. The thickness of the arrow represents the strength of the relationship when significant. All variables are included as independent observable variables. We grouped the different categories of predictors (soil properties, climatic legacies and spatial) in the same box in the model for graphical simplicity. For the same reason, we only included those direct effects from climatic legacies on soil properties that could indirectly affect the diversity of bacteria. The rest of the effects from climatic legacies on soil properties are available in Supplementary Tables 5 and 6. Acronyms for climatic and environmental variables are shown in Supplementary Tables 1 and 3, respectively. An 'a' adjacent to a particular chemical element indicates that the element is in an 'available' form. $R^2$, the proportion of variance explained. Significance levels of each predictor are $^*P < 0.05$, $^{**}P < 0.01$. Environmental drivers include plant diversity (Global Drylands, China and New South Wales) and/or soil properties and geographical location.

works on single response variables, we collapsed the bacterial community compositional data using non-metric multidimensional scaling (NMDS) for each dataset independently and retained the first two axes from a 2D solution (Bacterial community 1 and 2; stress ~0.1 in all cases). Prior to conducting SEM, we used a random forest[8] procedure (see Methods) to reduce the number of predictors to those that significantly explained the variation found in bacterial community richness and composition (geographical location, climatic legacies, soil properties and plant diversity) for each dataset (Supplementary Table 4). Random forest procedures are recommended for identifying the main significant predictors of environmental response variables (Methods). Finally, after conducting the random forest procedure but before the final SEM analyses, we ran preliminary SEMs to further evaluate whether the effects of climatic legacies on soil microbial community composition and richness were independent of those of current climate. We included in these analyses the selected climatic legacies from random forest analyses, but also included their corresponding current climate variables. In general, climatic legacies were as important as, or more important than, current climate in directly driving the richness and composition of bacteria across all datasets (Supplementary Fig. 4).

Our final SEM analyses provided solid evidence that climatic legacies had both direct effects (four of five cases) and indirect (four of five cases) effects on bacterial richness (Fig. 2; Supplementary Tables 5 and 6) across the five datasets used. Annual mean temperature (AMT) and precipitation in the driest month (PDM) showed the largest total effects (sum of direct and indirect effects from SEM) on bacterial richness in three of five datasets (Supplementary Fig. 5). Similarly, we also found both direct (four of five cases) and indirect (all cases) effects of climatic legacies on the composition of bacterial communities (for both NMDS axes; Supplementary Figs. 6-9; Supplementary Tables 5 and 6). In this case, direct effects were driven both by changes in temperature and in precipitation, with particular importance of AMT and isothermality, and of PDM (Supplementary Figs. 6 and 7). Indirect effects of precipitation and temperature legacies on microbial richness were largely driven by soil properties such as pH (three of five databases for bacterial richness and all databases cases for bacterial community composition), organic carbon concentration and texture (two of five cases for bacterial community composition, respectively; Fig. 2 and Supplementary Figs. 6 and 7). Other soil properties such as available phosphorus and micronutrients indirectly drove part of the effects of climatic legacies on microbial community structure (Fig. 2 and Supplementary Figs. 6 and 7). We also found strong indirect effects of climatic legacies on soil bacterial richness (three of three databases) through changes in plant diversity. In contrast, the effects of climatic legacies on bacterial community composition were indirectly driven by plant species richness only in China (Supplementary Fig. 6).

Additional random forest analyses (see Methods and Supplementary Data Table 7) allowed us to identify some of the bacterial taxa that were consistently (that is, in more than half of the datasets) good predictors of major climatic legacies (AMT and PDM, which were selected using standardized total effects from SEM; see Methods). For example, we found that the relative abundance of both *Planctomycetes* and candidate phylum WS3 (recently renamed *Latescibacteria*) consistently increased with increasing PDM from palaeoclimatic to current climates (Supplementary Data Table 7). In addition, phylum *Actinobacteria* was found to be an indicator of changes in temperature over millennia (Supplementary Data Table 7).

We repeated our variation partitioning models for a subset of data from the Australia dataset for which we were able to partition sites between croplands and natural ecosystems located close to these croplands (66 sites each). This allowed us to evaluate whether agriculture might alter the predictive power of palaeo- and current climates. We found that palaeoclimate (mid-Holocene plus Last Glacial Maximum) still predicted a unique part of the variation in bacterial diversity within croplands (Fig. 3 and Supplementary Fig. 10). However, palaeoclimate always had a significantly lower capacity to predict bacterial diversity in croplands than in natural ecosystems (Fig. 3 and Supplementary Fig. 10). When the SEMs were repeated using data from only natural and croplands sites in Australia (Fig. 3 and Supplementary Fig. 11), we found a strong reduction in the importance of soil properties as predictors of microbial community richness and composition, owing to the extreme disturbance caused by cotton and wheat farming (Fig. 3 and Supplementary Fig. 11).

## Discussion

Together, our work provides empirical evidence that palaeoclimate and climatic legacies (climate anomaly between 20,000 years ago and today) can leave a strong signature on soil bacterial communities, which may have influenced the contemporary distribution of bacterial richness and composition from regional to global scale. The importance of these results lies in the fact that climatic legacies can be used to better understand and predict the response of microbial communities to ongoing climate changes, including rising temperatures and changes in precipitation patterns[24]. For example, in arid environments (Global Drylands and New South Wales datasets; averages of 338/334 and 417/398 mm of current/Last Glacial Maximum annual precipitation, respectively), increasing PDM from palaeoclimates to current climates resulted in a net increase (sum of direct and indirect effects) in bacterial richness (Supplementary Fig. 5). This result is supported by a recent study highlighting that aridity, a proxy of water availability, is a key driver of bacterial diversity in global drylands[15]. However, in more humid environments such as those of the Americas and China (mostly temperate and tropical ecosystems; average of 948/894 mm and 903/1,020 mm of current/Last Glacial Maximum annual precipitation, respectively), increases in precipitation (Americas) or PDM (China) from palaeoclimate to current climates led to reductions in bacterial richness. This response is probably driven by increases in the relative abundance of specific microbial taxa under the wettest conditions[15] (Fig. 2 and Supplementary Fig. 5). For example, the relative abundance of both *Planctomycetes* and phylum *Latescibacteria* was positively related to the PDM anomaly from palaeoclimate to current climates (Supplementary Data Table 7). This result agrees with expectations that members of these phyla typically prefer wetter environments[25,26]. Interestingly, phylum *Actinobacteria* was also found to be an indicator of changes in temperature over millennia, indicating that this taxon may have been highly influenced by the last glaciation at the continental scale[5,6]. The effects of increasing precipitation on bacterial richness observed in China may be indirectly driven by soil acidification as a consequence of soil weathering[27] (Fig. 2d), as reductions in soil pH are known to reduce soil bacterial diversity in terrestrial ecosystems[13]. AMT showed the highest total (sum of direct and indirect effects) positive and negative effect on bacterial richness for the China and Australia datasets, respectively (Supplementary Fig. 5). This contrasting result might be related to the fact that Australia showed the lowest increase in temperature from palaeo- to current climates in this study (3.5 °C), which resulted in a total positive effect on the diversity of bacteria, compared with the increases found in China (5.6 °C), Global Drylands (5.2 °C) and the Americas (10.1 °C), where annual temperature legacies had a total negative effect on the diversity of soil bacterial communities.

Climatic legacies (measured as the temperature and precipitation anomalies[12] between an estimate of climate 20,000 years ago and another estimate for the present day) drove the richness and composition of bacterial communities both directly and indirectly through changes in soil properties and plant diversity. Direct

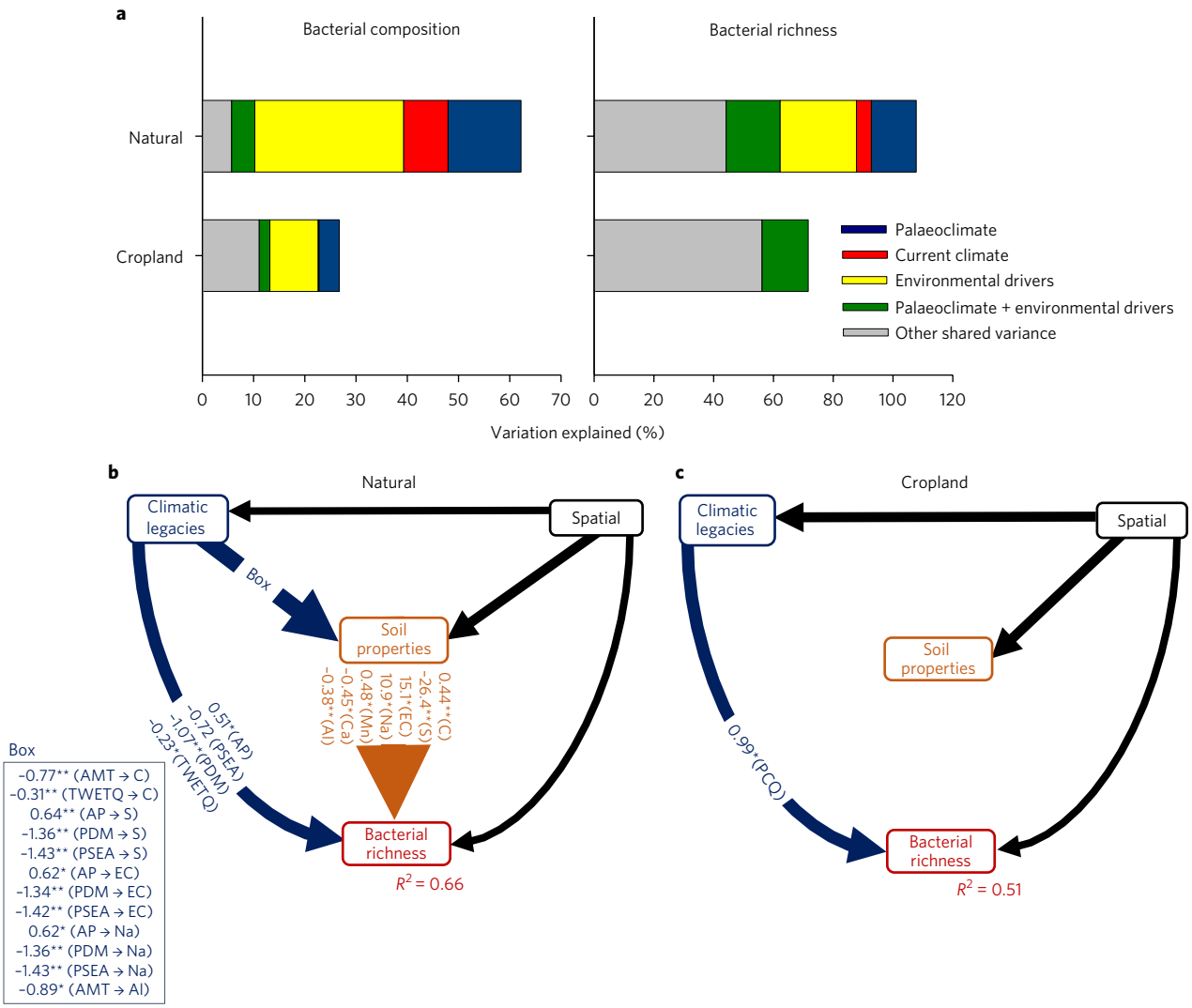


**Fig. 3 | Contribution of the different predictors for a subset of data from Australia. a**, Relative contribution of the different predictors used to model bacterial composition and diversity in croplands ($n = 66$) and natural ($n = 66$) ecosystems from Australia. Unique and shared variance from the Last Glacial Maximum and mid-Holocene in predicting bacterial community composition and richness were merged in this figure for simplicity. An alternative version of this figure showing the unique and shared variance of each group of predictors is available in Supplementary Fig. 10. **b**,**c**, Structural equation models accounting for the direct and indirect effects (plant diversity and/or soil properties) of climatic legacies on the diversity of soil bacteria in this database. Other details as in Fig. 2.

effects were driven both by changes in temperature and precipitation, with particular importance of AMT, isothermality and PDM (Supplementary Figs. 6 and 7). This finding is supported by recent studies that have identified temperature and precipitation as key global and continental-scale predictors of bacterial community richness and composition[15]. Direct effects include the impacts derived from rapid climatic changes in the past (mostly occurring prior to 10,000 years ago[12]) that have left a strong signature on the contemporary structure of soil bacterial communities (see Supplementary Appendix 3 for further rationale on direct effects from palaeoclimates on soil bacterial communities). Indirect effects of climatic legacies on bacterial richness and composition were largely driven by soil properties such as pH and, to a lesser extent, by organic carbon concentration and texture. These soil variables, which were included in all datasets, are known to determine changes in bacterial communities in terrestrial ecosystems[8,13–15]. Other soil properties such as soil phosphorus and micronutrients were also found to indirectly drive part of the effects of climatic legacies on bacterial community structure. In general, we found strong indirect effects of climatic legacies on soil bacterial richness via plant diversity, an indirect effect that was observed for bacterial composition only in China. Highly diverse plant communities may promote the richness of soil bacteria by supporting a wider variety of niches (for example, a range of litter qualities and rhizosphere products)[17]. Climatic legacies in these datasets always had direct and indirect (via soil properties) effects on plant richness, providing further support for the notion that such legacies play important roles in driving current plant diversity in terrestrial ecosystems[2–4].

As expected, geographical location, soil properties and plant richness accounted for significant variation in microbial community richness and composition in our model[8,13–15,18,28,29]. However, a unique and significant portion of variation was explained by climatic legacies, which suggests that geographical location, soil properties and plant diversity cannot account solely (via direct effects) for most of the effects of climatic legacies on soil bacterial communities (Figs. 1 and 2, and Supplementary Figs. 2, 3, 5–9). We acknowledge that other soil properties not included in our models

could improve the predictive power attributed to climatic legacies. Alternatively, part of the direct effect from climatic legacies on bacterial communities may still be indirectly driven by processes that we did not explicitly account for in the SEMs, such as dispersal-limited recolonization, which have been traditionally considered as main drivers of palaeoclimate effects on plant diversity[2–4].

Our analyses also provide evidence that palaeoclimate is influencing the contemporary distribution of bacterial communities in croplands. However, palaeoclimate always had a significantly lower capacity to predict the richness and composition of bacterial communities in croplands than in natural ecosystems. This suggests that agricultural practices have reduced the unique contribution of palaeoclimate as a predictor of contemporary soil microbial distribution. Our SEM results suggested that the influence of direct and indirect effects from climatic legacies on the richness and composition of soil bacterial communities is much lower in croplands (cotton and wheat farming) than in natural ecosystems. Agricultural practices can potentially erase or lessen the direct effects of such legacies on soil microbial communities by introducing new taxa of bacteria associated with the rhizosphere of particular crop species[19] and/or by artificially promoting particular bacterial species responsive to watering or fertilization[20]. Indirect effects from palaeoclimate on soil bacterial communities can be erased (richness) or lessened (composition) via rapid changes in soil carbon and pH as a result of agricultural practices. By radically changing the microbial communities in soil, agricultural practices hinder our capacity to predict the richness and composition of these bacterial communities using palaeoclimatic information. This result suggests that agriculture not only removes palaeoclimate legacies on microbial communities, but also leads to predictions with a lower level of accuracy than those obtained using data from natural ecosystems. Our results, coupled with those from previous studies[21], highlight the fact that agricultural intensification markedly alters soil microbial community composition and diversity in terrestrial ecosystems in ways that can be difficult to predict.

Together, our results provide strong evidence that past climates have left their signature on current bacterial diversity patterns across the globe and that agricultural practices may significantly reduce the unique signature of climatic legacies on soil bacteria. Overall, our findings indicate that using palaeoclimate data can improve our ability to predict the global distribution of soil bacterial communities in natural ecosystems. Thus, palaeoclimate data should be used when assessing the responses of these communities, and the ecosystem services that they provide, to global environmental change.

## Methods

**Study sites and data collection.** *Drylands (global scale).* We used data from ref. [15]. This dataset is focused on drylands (regions with an aridity index [precipitation/potential evapotranspiration] < 0.65)[30] and includes a wide variety of ecosystem types, including grasslands, shrublands and open woodlands, and environmental conditions across 'natural' dryland ecosystems. Field samples were collected between 2006 and 2010 from 80 sites located in 12 countries from all continents except Antarctica (Supplementary Fig. 1), under the most representative vegetation of each plot, according to a standardized sampling protocol[15]. A composite sample (from five soil samples; top 7.5 cm) was randomly taken under the canopy of the dominant perennial plant species. Each sample was separated into two portions. One portion was air-dried and used for chemical analyses. The other was immediately frozen at −20 °C for molecular analyses.

Soil DNA was extracted using the Powersoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. A portion of the bacterial 16S rRNA gene was sequenced using the Illumina MiSeq platform and the 341F/805R primer sets[31]. Bioinformatic analyses were conducted using the QIIME package[34] as described in ref. [15]. Operational taxonomic units (OTUs) were picked at 97% sequence similarity. The resultant OTU abundance tables from these analyses were rarefied to the same number of sequences per sample to ensure equal sampling depth (11,789). In addition, we removed OTUs that had only one read per OTU across all samples. Plant species richness and soil properties, including texture (percentage of sand), pH, electrical conductivity, total organic C, C:N ratio, total P, available P, available N (sum of inorganic and organic N), dissolved phenols, aromatic compounds, proteins, amino acids, carbohydrates and N mineralization, were measured as described elsewhere[32] (Supplementary Table 3).

*Americas (cross-continental scale).* We used data from ref. [28]. This dataset includes 48 'natural' sites across North and South America that cover a wide range of biomes and environmental conditions from Arctic to tropical forests (Supplementary Fig. 1). Composite soil samples (top 5 cm) were collected under the most representative vegetation of each study site. Each sample was separated into two portions. One portion was kept fresh and used for chemical analyses; the other was stored at −80 °C until DNA extraction.

Soil DNA was extracted using the Powersoil DNA Isolation Kit, following the manufacturer's instructions with the modifications described previously[13]. A portion of the bacterial 16S rRNA gene was sequenced using the Illumina MiSeq platform and the 515F/806R primer sets[33]. Bioinformatic analyses were conducted using the QIIME package[34] as described in ref. [28]. OTUs were picked at 97% sequence similarity. All samples were rarefied to 40,000 randomly selected reads per sample. In addition, we removed OTUs that had only one read per OTU across all samples.

Soil properties, including texture (percentage of sand), pH, total organic C, C:N ratio and C mineralization, were measured as described in ref. [13] (Supplementary Table 3).

*Australia (continental scale).* We used a subset of sample locations from the Biome of Australia Soil Environments (BASE) project[35] (Supplementary Fig. 1). This dataset include 531 soil samples belonging to 'natural' (465) and agricultural (66) (cropping by cotton and wheat) ecosystems from Australia. Samples were collected between 2011 and 2014. Soil samples were collected according to the methods described in ref. [35]. In brief, at each plot, a composite soil sample (nine discrete soil samples) from the top 0–10 cm was collected and separated into two portions[35]. One portion was air-dried for chemical analyses, the other was frozen (−80 °C) until DNA extraction.

All soil DNA was extracted in triplicate, according to the methods used by the Earth Microbiome Project[35], at the Australian Genome Research Facility. Amplicons targeting the bacterial 16S rRNA gene were sequenced using the Illumina MiSeq platform and the 27F – 519R[36] primer set (see ref. [35] for details on these analyses). Bioinformatic analyses were performed using the mothur software (v1.34.1)[37] as explained in ref. [35]. OTUs were picked at 97% sequence similarity. The OTU abundance tables were rarefied at 14,237 sequences per sample to ensure even sampling depth. In addition, we removed OTUs that had only one read per OTU across all samples.

Soil properties including texture (percentage of sand), pH, electrical conductivity, total organic C, available N (sum of ammonium and nitrate), available P, available K, and total K, S, Cu, Fe, Mn, Zn, Al, Ca, Mg, Na and B were measured as described in ref. [35] (Supplementary Table 3).

*China (continental scale).* This dataset focuses on forest ecosystems (boreal, temperate mixed coniferous, temperate deciduous, subtropical evergreen and tropical forests) and includes 300 plots across a wide latitudinal gradient (approximately 4,000 km) in Eastern China[29] (Supplementary Fig. 1). In each plot, a composite soil sample from 15 soil cores was collected from the top 0–10 cm. Each sample was separated into two portions. One portion was air-dried and used for soil chemical analyses, and the other was stored at −80 °C until DNA extraction.

Soil DNA was extracted using the Powersoil DNA Isolation Kit, with a slight modification as explained in ref. [29]. A portion of the 16S rRNA gene (515F/806R primer set)[37] was sequenced using the Illumina MiSeq platform. Bioinformatic analyses were completed using the QIIME pipeline[34] (see ref. [29] for details on these analyses). OTUs were picked at 97% sequence similarity. All samples were rarefied to 40,000 randomly selected reads per sample. In addition, we removed OTUs that had only one read per OTU across all samples.

Plant species richness and soil properties, including texture (percentage of sand), pH, total organic C, C:N ratio and available P, were measured as described in ref. [29] (Supplementary Table 3).

*New South Wales (regional scale).* We used data from ref. [38]. This dataset includes data from 54 sites scattered across a 500-km$^2$ area of semi-arid eastern Australia (Supplementary Fig. 1). This survey was undertaken in three woodland communities dominated by blackbox (*Eucalyptus largiflorens*), white cypress pine (*Callitris glaucophylla*) and river red gum (*Eucalyptus camaldulensis*). This dataset includes sites extensively used for livestock grazing, large areas dedicated for conservation (national parks, nature reserves) and smaller areas devoted to native forestry. At each site, a soil sample was collected in 2014 from the top 5 cm of soil. For this study, we used the subset of samples collected under tree microsites[38]. Each sample was separated into two portions. One portion was air-dried and used for soil chemical analyses; the other was stored at −20 °C until DNA extraction.

Soil DNA was extracted using the Powersoil DNA Isolation Kit according to the manufacturer's instructions. Amplicons targeting the bacterial 16S rRNA gene were sequenced using the Illumina MiSeq platform and the 341F/805R primer set[31]. Bioinformatic analyses were done using mothur[37] and UPARSE[39]. OTUs were picked at 97% sequence similarity. We removed OTUs that had only

one read per OTU across all samples, and the resulting OTU abundance tables were rarefied to 10,851 sequences per sample (the fewest sequences obtained in a single soil sample).

At each of three positions along a 100-m transect (0 m, 50 m, 100 m), we selected the nearest tree (perennial plant >4 m in height). Two small (0.5 m × 0.5 m) quadrats were placed midway between the trunk and the canopy edge on opposite sides of the canopy. Within these small quadrats, we assessed the cover of all vascular plants by species and used these data to obtain a value of total plant species richness for each site. Soil properties including texture (percentage of sand), bulk density and available P were measured as described in ref. [38] (Supplementary Table 3).

**Climate data.** In all cases, a total of 19 standardized climatic variables (Supplementary Table 1) were obtained for all the sites surveyed from the Worldclim database (www.worldclim.org). In the case of mid-Holocene and Last Glacial Maximum climates, we used estimates provided by the Community Climate System Model (CCSM4; www.worldclim.org)[40–43]. We used data at a spatial resolution of 2.5 minutes (~4.5 km at the Equator), as this is the highest resolution available for the Last Glacial Maximum period. Bioclimatic data are also available for this resolution for current and mid-Holocene climates, allowing direct comparison among bioclimatic data at different periods. In all cases, climate data were at a spatial resolution of 30 seconds for current and mid-Holocene, which allowed us to compare data at resolutions of 2.5 minutes and 30 seconds for these two periods. Values calculated using 2.5 minutes were identical to those calculated using a resolution of 30 seconds in all cases (Pearson's $r > 0.99$; $P < 0.001$) (see 'Climatic data cross-validation' in Supplementary Appendix 1). We acknowledge that palaeoclimatic data from islands may be inaccurate as a consequence of their spatial location, which influences the accuracy of the available climate data. This should not, however, bias the conclusions from this study, which was conducted at the global scale, given the low number of data points coming from islands[12].

**Pre-selection of multicollinearity free climatic variables.** We decided to use only those climate variables (from the original 19 climate variables available from Worldclim) that were free of multicollinearity within each period of time (that is, current, Last Glacial Maximum and mid-Holocene). For example, the inclusion of strongly positively correlated ($r > 0.8$) variables[44] within a particular group of predictors is not recommended for variation partitioning modelling as they may cause multicollinearity problems in the analyses (see below). To preselect multicollinearity-free climate variables from the original list, we collapsed the climate information from all datasets and conducted correlation analyses (Pearson) within each period of time (that is, group of predictors in our variation partitioning) for the original 19 climate variables available (Supplementary Tables 1 and 2). Based on these analyses, we selected for our analyses the same 8 out of 19 climatic variables for each period of time that were not strongly correlated with the rest ($r < 0.8$)[44]: annual mean temperature (AMT), mean diurnal temperature range (MDR), isothermality (ISO), temperature in the wettest quarter (TWETQ), annual precipitation (AP), precipitation in the driest month (PDM), precipitation seasonality (PSEA) and precipitation in the colder quarter (PCQ). These eight variables include variables that are highly correlated to multiple non-selected variables and also variables that were unrelated to any other climate variable, hence could only be explained by themselves. Together, these variables are a good representation of the rest of non-selected climatic variables (Supplementary Table 2). We retained these eight variables for the remainder of statistical analyses presented in this manuscript.

**Statistical modelling.** We used a combined approach including multiple statistical models to address our different hypotheses. In particular, we used (1) variation partitioning modelling to identify whether palaeoclimate can explain a unique portion of the variation in bacterial community richness and composition that cannot be accounted for other key predictors of soil microbial communities; (2) random forest analysis to identify the main individual predictors of bacterial community richness and composition including spatial predictors, climatic legacies, soil properties and plant richness; and (3) SEM to identify the direct and indirect (via soil properties and plant richness) effects of climatic legacies on bacterial community richness and composition. All of these statistical models address a particular part of our research question that cannot be addressed using each approach on its own.

**Variation partitioning modelling.** We used variation partitioning[23] to quantify the relative importance of four groups of predictors: (1) climate variables from the Last Glacial Maximum, (2) climate variables from the mid-Holocene, (3) climate variables from current climates and (4) other key environmental drivers of microbial communities, including plant species richness (available for Global Drylands, China and New South Wales surveys) and/or soil properties (14 for Global Drylands, 5 for Americas, 18 for Australia, 5 for the China and 6 for the New South Wales survey; total organic carbon, texture and pH were included in all models) and space (site location as defined by latitude and longitude) as predictors of the bacterial community composition (number of reads per OTU) and richness (that is, number of OTUs per sample). Climate includes the eight multicollinearity-free variables described above (Supplementary Tables 1 and 2). Geographical location (latitude and longitude) was included in all models to account for spatial autocorrelation (see ref. [15] for a similar approach). The complete list of predictors available for each database is presented in Supplementary Table 3.

Variation partitioning is a method specifically recommended to deal with between-group multicollinearity, as it partitions the variance in a given response variable that is attributed to a particular group of predictors from that variance shared among all predictors[23]. Thus, this analysis allows us to identify whether climate variables from current, mid-Holocene and Last Glacial Maximum periods can explain a unique portion of the variance that is not explained by climate in other periods[23]. Note that adjusted coefficients of determination in multiple regression and canonical analysis can, on occasion, take negative values[23]. Negative values in the variance explained for a group of predictors on a group of response variable are interpreted as zeros and correspond to cases in which the explanatory variables explain less variation than random normal variables would[23]. In all cases, variation partitioning analyses were conducted with the R package Vegan[45]. The complete list of predictors available for each database is presented in Supplementary Table 3.

*Assessing comprehensive indices of climatic legacies.* To obtain a greater mechanistic understanding of the role of palaeoclimate in regulating current microbial richness and composition, we calculated specific climatic legacy indices for each of the preselected eight climatic variables. To do so, we calculated the mathematical difference in the values for each climatic variable from Last Glacial Maximum to current climates (annual $\text{precipitation}_{\text{Current climate}}$ − annual $\text{precipitation}_{\text{Last Glacial Maximum}}$) for each site. Therefore, climatic legacies represent the temperature and precipitation anomalies between an estimate of climate 20,000 years ago and another estimate for the present day[12]. This difference informs us about the climatic legacies—increases, decreases or lack of changes for a particular climate condition with time—in each of the sites surveyed from the different datasets (see 'Climatic legacy indexes cross-validation' in Supplementary Appendix 2). We used palaeoclimatic information from the Last Glacial Maximum rather than mid-Holocene conditions because (1) the former is included in the period between Last Glacial Maximum and the current climate, and (2) in general, Last Glacial Maximum conditions were a better predictor of bacterial richness and composition than climatic conditions in the mid-Holocene (Supplementary Figs. 2 and 3; Supplementary Appendix 2). Note that the climatic legacy index used here is based on the differences between two single snapshots in time (current versus Last Glacial Maximum climates), and thus calculation of climatic legacy comes with several inherent and important assumptions[12], some of which are accounted for in Supplementary Appendix 2. Also, most abrupt changes in climate occurred before 10,000 years ago[12] (see, for example, Supplementary Figs. 12–14). Even so, our climatic legacy index still allowed us to address our research question on whether the difference between climate today and 20,000 years ago affects the structure of current bacterial communities, and whether these effects were directly mediated by climatic legacies or indirectly through soil properties and plant diversity.

*Random forest modelling I: pre-selection of main microbial drivers used in structural equation modelling.* Owing to the large number of predictors used, we conducted a classification random forest analysis[46] as described in ref. [8] to identify the main statistically significant predictors of the composition and richness of bacteria to be included in our structural equation models (next section). Unlike the variation partitioning model described above, both random forest and structural equation modelling take one response variable at each time. In this respect, in the case of bacterial community composition at the OTU level, we conducted random forest analyses on the two scores from the 2D solution of a NMDS using the Bray–Curtis dissimilarity metric (Bacterial community 1 and 2). The complete list of predictors available for each database is presented in Supplementary Table 3. These analyses were conducted using the rfPermute package[47] of the R statistical software (http://cran.r-project.org/).

*Structural equation modelling.* We used SEM[48] to evaluate the direct effects (changes in temperature and precipitation variables with time) and indirect effects (plant diversity and/or soil properties and space) of climatic legacies on the richness and composition of bacterial communities. The use of SEM is particularly useful in large-scale correlative studies, as it allows the partitioning of causal influences among multiple variables, and separation of the direct and indirect effects of model predictors[48]. The main structure of our a priori model was shared across all datasets and response variables (Supplementary Fig. 15). We only included in these models those variables that were identified as major statistically significant predictors of the richness and composition of bacterial communities from random forest analyses (Supplementary Table 4). Therefore, SEM models conducted for the different datasets contain different predictors and were independently constructed. The only exceptions to this were latitude and longitude, which were included in all the models to account for spatial autocorrelation in our models (as done in ref. [15]). By simplifying our models with such an approach, we acknowledge that we may be missing some indirect effects from excluded variables on bacterial community richness or

composition. However, we also reduce the complexity of our models, providing a more comprehensive understanding on the main direct and indirect effects from climatic legacies on the richness and composition of bacterial communities, therefore allowing us to properly address our research question. A consequence of this approach is that direct effects between climatic legacies and soil properties may not include some of the major climatic legacies controlling soil properties, obscuring the interpretation of these parts of the models. It was not the goal of this study to identify the major direct effects of climatic legacies on soil properties. Consequently, we only included in our models those soil properties that directly influenced soil bacterial community richness or composition, or that ultimately could lead to indirect effects of climatic legacy on bacterial communities.

In our models, all variables are included as independent observed variables. We grouped the different categories of predictors (climatic legacies, soil properties and spatial) in the same box in the model for graphical simplicity, but these boxes do not represent latent variables. The climatic legacies box includes all selected individual climatic legacies identified as significant predictors of bacterial community richness or composition from random forest analyses. The spatial box includes latitude and longitude. The soil properties box include all individual soil properties identified as significant predictors of bacterial community richness or composition by our random forest analyses. Note that if none of the variables within a particular box (for example plant richness or soil properties) was selected by our random forest analyses as significant predictors of a particular microbial variables and for a particular dataset, that box is excluded in that specific model. We included both direct and indirect (via soil properties and plant richness) effects of climatic legacies on the richness and composition of bacterial communities in our models (see Supplementary Appendix 3).

All variables within a particular box were allowed to covary. Because of this, all models were originally saturated (zero degrees of freedom). To release a degree of freedom and make it possible for us to test the goodness of fit of our models, we conducted the following a priori analyses: (1) we conducted partial correlations (Pearson) between all predictors within a particular model and (2) we removed the weakest a priori correlation (Supplementary Table 9) between two predictors in our models. The goodness of fit of SEM models was checked following ref. [49]. 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[49]. We used five measures of goodness of fit of our models including the (1) comparative fit index (CFI) (the model has a good fit when $0.97 \leq \mathrm{CFI} \leq 1.00$ and acceptable fit when $0.95 \leq \mathrm{CFI} < 0.97$); (2) goodness-of-fit index (GFI) (the model has a good fit when $0.95 \leq \mathrm{GFI} \leq 1.00$ and acceptable fit when $0.90 \leq \mathrm{GFI} < 0.95$); (3) normed fit index (NFI) (the model has a good fit when $0.95 \leq \mathrm{NFI} \leq 1.00$ and acceptable fit when $0.90 \leq \mathrm{NFI} < 0.95$); (4) $\chi^2$ test ($\chi^2$; the model has a good fit when $0 \leq \chi^2/\mathrm{d.f.} \leq 2$ and $0.05 < P \leq 1.00$, and acceptable fit when $2 < \chi^2/\mathrm{d.f.} \leq 3$ and $0.01 \leq P \leq 0.05$); and (5) the root mean square error of approximation (RMSEA; the model has a good fit when $0 \leq \mathrm{RMSEA} \leq 0.05$ and $0.10 < P \leq 1.00$, and acceptable fit when $0.05 < \mathrm{RMSEA} \leq 0.08$ and $0.05 \leq P \leq 0.10$)[49]. In general, our a priori models attained an acceptable or good fit. In particular, 16/21 cases showed a good/acceptable fit by all criteria (Supplementary Table 10). The remaining 5/21 cases still showed a good fit in three of the five indexes used here (CFI, GFI and NFI) (Supplementary Table 10). No post-hoc alterations were made. With an acceptable/good model fit, we were free to interpret the path coefficients of the model and their associated $P$ values. SEM models were conducted with the software AMOS 20 (IBM SPSS Inc., Chicago, IL, USA).

We also calculated the standardized total effects of plant diversity and/or soil properties, space and climatic legacies on the richness and composition of bacteria. The net influence that one variable has on another is calculated by summing all direct and indirect pathways between the two variables. If the SEM model fits the data well, the total effect should approximate the bivariate correlation coefficient for that pair of variables[48].

*Random forest modelling II: identifying the main phyla characterizing particular climatic legacies.* We also used random forest analysis[46] as described in ref. [8] to identify the main bacterial phyla predicting a particular climatic legacy. We focused on the main climatic legacies driving bacterial community composition, which were identified using the standardized total effects from SEM: AMT (in China and Australia) and PDM (in Global Drylands and New South Wales) (Supplementary Figs. 8 and 9). AMT was also a major predictor of the bacterial community composition in the Americas dataset (Supplementary Fig. 8). In these analyses, the relative abundance of bacterial phyla acts as a predictor of a particular legacy variable (AMT or PDM). These phyla include: *Thermi*, *Acidobacteria*, *Actinobacteria*, *AD3*, *Armatimonadetes*, *Bacteroidetes*, *BRC1*, *Chlorobi*, *Chloroflexi*, *Cyanobacteria*, *Elusimicrobia*, *FBP*, *FCPU426*, *Fibrobacteres*, *Firmicutes*, *Gemmatimonadetes*, *GN02*, *Nitrospirae*, *NKB19*, *OD1*, *OP11*, *OP3*, *Planctomycetes*, *Proteobacteria*, *Tenericutes*, *TM6*, *TM7*, *Verrucomicrobia*, *WPS-2*, *WS2*, *WS3* and *WS4*. These analyses were conducted using the rfPermute package[47] of the R statistical software (http://cran.r-project.org/). The main goal of these analyses is to identify the main taxa characterizing a particular climatic legacy from AMT or PDM.

We first identified the main (that is, significant; $P < 0.05$, according to random forest results) microbial phyla accounting for the variation of particular climatic legacies and that are highly related to a particular legacy in a consistent way (those microbial taxa that are selected from a random forest model as important drivers of either AMT or PDM in more than half of the datasets; Supplementary Table 7). We then identified the shape of the relationship between climatic legacies and the relative abundance of selected taxa. All statistical analyses were independently performed with each dataset. To identify the best shape describing the relationship between climatic legacies and microbial taxa, we fitted two different functions that involve different biological interpretations: linear (positively or negatively affected by precipitation and temperature legacies) and quadratic (microbial taxa that are positively or negatively affected by intermediate levels of precipitation and temperature legacies). We selected the best model fits by following Akaike information criteria (AIC)[50]. The lower the AIC index, the better the model. Here, we consider a $\Delta\mathrm{AIC} > 2$ threshold to differentiate between two different models and then select the best of those models (see ref. [50] for a similar approach). When both quadratic and linear models were similar ($\Delta\mathrm{AIC} < 2$), we selected the linear (most parsimonious) model.

**Data accessibility.** Data associated with this paper have been deposited in figshare: https://figshare.com/s/e3e47dc51ac2090f38bb (DOI: 10.6084/m9.figshare.5048311).

Received: 5 February 2017; Accepted: 28 June 2017;
Published online: 7 August 2017

## References

1. Monger, C. et al. Legacy effects in linked ecological–soil–geomorphic systems of drylands. *Front. Ecol. Environ.* **13**, 13–19 (2015).
2. Gajewski, K. Impact of Holocene climate variability on Arctic vegetation. *Glob. Planet. Chang.* **133**, 272–287 (2015).
3. Svenning, J.-C. et al. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annu. Rev. Ecol. Evol.* **46**, 551–572 (2015).
4. Lyons, S. K. et al. Holocene shifts in the assembly of plant and animal communities implicate human impacts. *Nature* **529**, 80–83 (2016).
5. Martiny, J. B. H. History leaves its mark on soil bacterial diversity. *mBio* **7**, e00784-16 (2016).
6. Andam, C. P. et al. A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *mBio* **7**, e02200-15 (2016).
7. Wagg, C. et al. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proc. Natl Acad. Sci. USA* **111**, 5266–5270 (2014).
8. Delgado-Baquerizo, M. et al. Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* **28**, 10541 (2016).
9. van der Heijden, M. G. et al. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310 (2008).
10. Oliverio, A. et al. Identifying the microbial taxa that consistently respond to soil warming across time and space. *Glob. Chang. Biol.* **23**, 2117–2129 (2017).
11. Atkinson, T. C. et al. Seasonal temperatures in Britain during the past 22,000 years, reconstructed using beetle remains. *Nature* **325**, 587–592 (1987).
12. Fordham, D. A. et al. PaleoView: a tool for generating continuous climate projections spanning the last 21000 years at regional and global scales. *Ecography* http://dx.doi.org/10.1111/ecog.03031 (2017).
13. Lauber, C. L. et al. Soil pH as a predictor of soil bacterial community structure at the continental scale: a pyrosequencing-based assessment. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
14. Delgado-Baquerizo, M. et al. Carbon content and climate variability drive global soil bacterial diversity patterns. *Ecol. Monogr.* **86**, 373–390 (2016).
15. Maestre, F. T. et al. Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proc. Natl Acad. Sci. USA* **112**, 15684–15689 (2015).
16. Vitousek, P. M. *Nutrient Cycling and Limitation: Hawai'i as a Model System* (Princeton University, New Jersey, 2004).
17. Wardle, D. A. et al. Ecosystem properties and forest decline in contrasting long-term chronosequences. *Science* **305**, 509–513 (2004).
18. Prober, S. M. et al. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecol. Lett.* **18**, 85–95 (2015).
19. Leff, J. W. et al. Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, *Helianthus annuus*. *New Phytol.* **214**, 412–423 (2017).
20. Leff, J. W. et al. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc. Natl Acad. Sci. USA* **112**, 10967–10972 (2015).
21. Trivedi, P. et al. Response of soil properties and microbial communities to agriculture: implications for primary productivity and soil health indicators. *Front. Plant Sci.* **7**, 990 (2016).
22. *World Development Report: Agriculture for Development* (World Bank, Washington DC, 2008).
23. Legendre, P. et al. Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *J. Plant Ecol.* **1**, 3–8 (2008).

24. Nogués-Bravo, D. et al. Amplified plant turnover in response to climate change forecast by Late Quaternary records. *Nat. Clim. Chang.* **6**, 1115–1119 (2016).
25. Fieseler, L. et al. Discovery of the novel candidate phylum 'Poribacteria' in marine sponges. *Appl. Environ. Microbiol.* **70**, 3724–3732 (2004).
26. Youssef, N. H. et al. *In silico* analysis of the metabolic potential and niche specialization of candidate phylum '*Latescibacteria*' (WS3). *PLoS ONE* **10**, e0127499 (2015).
27. Schlesinger, W. H. & Bernhardt, E. S. *Biogeochemistry: an Analysis of Global Change* (Academic Press, San Diego, 1996).
28. Ramirez, K. S. et al. Biogeographic patterns in belowground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. Lond. B Biol. Sci.* **281**, 20141988 (2014).
29. Wang, J.-T. et al. Coupling of soil prokaryotic diversity and plant diversity across latitudinal forest ecosystems. *Sci. Rep.* **5**, 19561 (2015).
30. *United Nations Environment Programme World Atlas of Desertification* (Edward Arnold, London, 2002).
31. Herlemann, D. P. et al. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**, 1571–1579 (2011).
32. Maestre, F. T. et al. Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335**, 214–218 (2012).
33. Bates, S. et al. Bacterial communities associated with the lichen symbiosis. *Appl. Environ. Microbiol.* **77**, 1309–1314 (2011).
34. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
35. Bissett, A. et al. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience* **20165**, 21 (2016).
36. Lane, D. J. *Nucleic Acid Techniques in Bacterial Systematics* (Wiley, New York, 1991).
37. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol* **75**, 7537–7541 (2009).
38. Eldridge, D. J. et al. Competition drives the response of soil microbial diversity to increased grazing by vertebrate herbivores. *Ecology* **98**, 1922–1931 (2017).
39. Edgar, R. G. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
40. Gent, P. R. et al. The Community Climate System Model version 4. *J. Clim.* **24**, 4973–4991 (2011).
41. Bystriakova, N. et al. Present, past and future of the European rock fern *Asplenium fontanum*: combining distribution modelling and population genetics to study the effect of climate change on geographic range and genetic diversity. *Ann. Bot.* **13**, 453–465 (2013).
42. Tallavaaraa, M. et al. Human population dynamics in Europe over the Last Glacial Maximum. *Proc. Natl Acad. Sci. USA* **112**, 8232–8237 (2015).
43. Delgado-Baquerizo, M. et al. Biogeographic bases for a shift in crop C:N:P stoichiometries during domestication. *Ecol. Lett.* **19**, 564–575 (2016).
44. Katz, M. H. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers* (Cambridge Univ. Press, Cambridge, 2006).
45. Oksanen, J. et al. vegan: Community Ecology Package. R package v. 2.3-0 (2015).
46. Breiman, L. Random forests. *Mach. Learn.* **45**, 5 (2001).
47. Archer, E. rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics. R package v. 1.5.2 (2016).
48. Grace, J. B. *Structural Equation Modeling Natural Systems* (Cambridge Univ. Press, Cambridge, 2006).
49. Schermelleh-Engel, K. et al. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* **8**, 23–74 (2003).
50. Delgado-Baquerizo, M. et al. Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. *J. Ecol.* **104**, 936–946 (2016).


## Acknowledgements

M.D.-B. acknowledges support from the Marie Sklodowska-Curie Actions of the Horizon 2020 Framework Programme H2020-MSCA-IF-2016 under REA grant agreement no. 702057. We acknowledge the contribution of the BASE project partners (DOI: 10.4227/71/561c9bc670099), an initiative supported by Bioplatforms Australia with funds provided by the Australian Commonwealth Government through the National Collaborative Research Infrastructure Strategy. B.K.S. and M.D-B are supported by the Australian Research Council projects (DP13010484 and DP170104634). D.J.E. was supported by the Hermon Slade Foundation. N.F was supported by grants from the US National Science Foundation (PLR 1241629 and DEB 1542653). The work from J.-Z.H. and the China dataset were supported by the Natural Science Foundation of China (grant no. 41230857) and the Chinese Academy of Sciences (grant no. XDB15020200). The work of F.T.M. and the Global Drylands database were supported by the European Research Council (ERC grant agreements 242658 [BIOCOM] and 647038 [BIODESERT]).

## Author contributions

M.D.-B. conceived the idea of this study in consultation with N.F. The microbial datasets of the Global Drylands were originally compiled by F.T.M, B.K.S. and M.D.-B.; those of the Americas by N.F.; Australia by A.B.; China by J.-Z.H., Y.-R.L. and J.-T.W.; and New South Wales by D.J.E., B.K.S. and K.H. Statistical modelling was conducted by M.D.-B. The manuscript was written by M.D.-B. with contributions from all co-authors.

## Competing interests

The authors declare no competing financial interests.

## Additional information

**Supplementary information** is available for this paper at doi:10.1038/s41559-017-0259-7.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to M.D.-B.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.