Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2014) **23**, 1346–1355

A Journal of Macroecology

Global Ecology and Biogeography

META-ANALYSIS

A role for biotic filtering in driving phylogenetic clustering in soil bacterial communities

Marta Goberna[1,2], Carlos García[1] and Miguel Verdú[2]*

[1]*Centro de Edafología y Biología Aplicada del Segura (CEBAS-CSIC), Campus Universitario de Espinardo, E-30100, Espinardo, Murcia, Spain,* [2]*Centro de Investigaciones sobre Desertificación (CSIC-UVEG-GV), Carretera Moncada-Náquera, Km 4.5. E-46113, Moncada, Valencia, Spain*

*Correspondence: Miguel Verdú, Centro de Investigaciones sobre Desertificación (CSIC-UVEG-GV), Carretera Moncada-Náquera, Km 4.5. E-46113, Moncada, Valencia, Spain.
E-mail: miguel.verdu@uv.es

ABSTRACT

Aim Phylogenetic clustering, the coexistence of evolutionarily related organisms, appears to be common in soil bacteria. This pattern has traditionally been attributed to the habitat-filtering of bacteria that are able to survive under particular abiotic settings. According to the modern coexistence theory, however, phylogenetic clustering can also arise from biotic interactions such as the competitive exclusion of large clades with low competitive abilities. Here, we used phylogeny-based methods to discern whether the coexistence of evolutionarily related soil bacteria results from abiotic and/or biotic filtering.

Location Worldwide.

Methods We performed a Bayesian meta-analysis based on a literature review ($n = 231$) to assess whether the net relatedness index (*NRI*) or the nearest taxon index (*NTI*), two measures of the phylogenetic relatedness of taxa in local assemblages, deviate from those in randomly configured communities. We then sought the best abiotic (pH, total organic carbon and total nitrogen) and biotic predictors (relative abundance of *Proteobacteria*, *Actinobacteria* and *Acidobacteria*) of *NRI* and *NTI*.

Results Phylogenetic clustering is pervasive in soil bacterial communities regardless of the spatial and taxonomic scales ($NRI = 2.29$; 95% CI [1.43, 3.29]; $P < 0.001$). Clustering is accentuated by productivity; that is, more fertile soils hold communities with more closely related bacteria (estimate = 1.05 [0.03, 2.15]; $P < 0.05$). Proteobacterial abundance, which increases with organic carbon enrichment, leads to higher relatedness among coexisting bacteria (estimate = 0.1 [0.02, 0.17]; $P < 0.01$) through the competitive exclusion of distantly related deep-branching clades.

Main conclusions Our results, together with the dominance of proteobacterial lineages in soils worldwide, suggest that the overrepresentation of this clade underlies the widespread coexistence of phylogenetically related bacteria. These results are consistent with phylogenetic clustering arising via differences in competitive ability as predicted by the coexistence theory. This supports the idea that biotic filtering might have a role in driving the phylogenetic community assembly of soil prokaryotes.

Keywords
Community structure, competitive exclusion, environmental filtering, phylogenetic diversity, *Proteobacteria*, soil organic carbon.

INTRODUCTION

The processes that underlie community assembly in bacteria have long enticed microbiologists (Baas Becking, 1934). A common view is that bacteria have cosmopolitan distributions, and the local abiotic scenario determines the species composition of ecological communities (Baas Becking, 1934; Horner-Devine *et al.*, 2004; Fierer & Jackson, 2006; Martiny *et al.*, 2006). Soil bacteria are crucial components of the biosphere because they are directly related to the soil's fertility, and are ultimately linked to plant productivity and diversity (van der Heijden *et al.*, 2008). Co-occurrence patterns of soil bacteria have been attributed to

DOI: 10.1111/geb.12227

habitat sorting through abiotic factors such as pH, organic carbon or mineral nitrogen (Fierer & Jackson, 2006; Smith *et al.*, 2008; Goberna *et al.*, 2012). Here, we argue that biotic factors (competition) might be more influential in shaping soil bacterial communities than contemporary literature depicts.

The incorporation of phylogenetic information into the analysis of co-occurrence patterns was initially designed to infer the abiotic and biotic processes that structure communities (Webb *et al.*, 2002). Community phylogenetics uses phylogenetic relatedness among taxa as an indicator of their ecological similarity (Pausas & Verdú, 2010). This relies on the phylogenetic conservatism of functional traits, an assumption that holds true for prokaryotes, as Martiny *et al.* (2013) demonstrated through analysis of the phylogenetic distribution of 89 metabolic traits using both genetic data (over 2220 genomes) and phenotypic data (organic C consumption patterns of 738 strains). Phylogeny-based methods compare the relatedness of taxa in a local community to those of randomly configured communities assembled from the regional taxon pool (Webb *et al.*, 2002). Phylogenetic clustering, or the co-occurrence of closely related taxa more often than expected by chance, seems to prevail in soil bacteria (Horner-Devine & Bohannan, 2006; Lozupone & Knight, 2007; Bryant *et al.*, 2008; Jones *et al.*, 2009), although overdispersion and randomness have also been detected (Horner-Devine & Bohannan, 2006; Costello *et al.*, 2009; Chong *et al.*, 2012). Several authors have posited that a dominance of phylogenetic clustering in bacterial communities might emerge from working at broad spatial and/or taxonomic scales (Horner-Devine & Bohannan, 2006; Bryant *et al.*, 2008). This is due to the increasing probability of detecting clustering as the sampling area is enlarged, because the randomly built communities will be more likely to include high-level clades (e.g. phyla) that are absent from the local assemblage (Swenson *et al.*, 2006). The relevance of this process depends, however, on the rate at which new taxa are sampled as the area is scaled out. Bacteria have the lowest magnitude of change in the taxa–area relationship among all major groups of organisms (Horner-Devine *et al.*, 2004), and the extent to which phylogenetic community patterns of soil bacteria are scale-dependent therefore remains to be elucidated.

Phylogenetic clustering has been traditionally interpreted as the result of the environmental filtering of organisms that bear phylogenetically conserved traits (Webb *et al.*, 2002). The habitat restricts the survival in a community to ecologically (and phylogenetically) similar organisms that are able to thrive under a specific abiotic setting. Organisms that are able to surpass such a filter tolerate abiotic factors that determine their survival and adaptation to the environment (e.g. pH, desiccation, radiation and salinity) (Webb *et al.*, 2002). Abiotic filtering is illustrated in our hypothetical example in Fig. 1a, in which only the bacteria with resistant cell walls (represented by contour thickness) are not filtered out at high levels of radiation. In this example, tolerance to radiation is phylogenetically conserved, and so the members of the abiotically filtered communities are more closely related than expected at random. Alternatively, Mayfield & Levine (2010) have recently proposed, based on modern coexistence theory (Chesson, 2000), that biotic interactions (i.e. competition) can also generate phylogenetically clustered patterns. This occurs when competitive exclusion takes place among distantly related organisms – that is to say, when a few phylogenetic clades outcompete all others given their superior competitive ability under the prevailing abiotic environment. Such a biotic filter implies that certain organisms consume limiting resources (e.g. organic substances or mineral ions) that are no longer available to others. This competitive interaction results in the simultaneous outgrowth and suppression of the

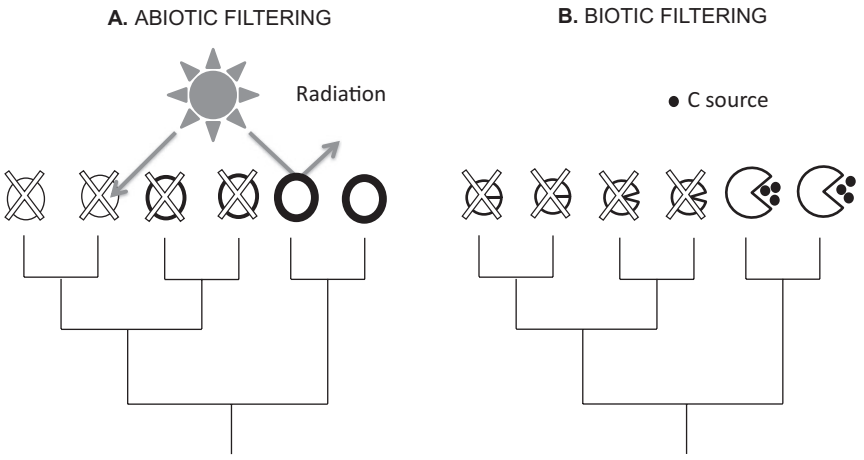

Figure 1 Phylogenetic clustering can be generated through abiotic or biotic filtering. In example (a), bacteria differ primarily in their tolerance to radiation (contour thickness) and this trait is phylogenetically conserved. At high levels of radiation, only tolerant bacteria survive the filter, generating phylogenetically clustered communities. In example (b), bacteria differ primarily in their growth response to carbon substrates (cell size) and this trait is phylogenetically conserved. In the presence of a carbon source, fast-growers outcompete distantly related clades, generating phylogenetically clustered communities. Note that organisms that survive an abiotic filter (a) show tolerance to environmental conditions, whereas those surviving a biotic filter (b) have the ability to exploit limiting resources.

strong and weak competitors, respectively. Biotic filtering is exemplified in Fig. 1b, in which fast-growing bacteria (represented by cell size) stimulated by a carbon source outcompete distantly related clades. In the example, growth rate in response to the substrate is phylogenetically conserved; thus, the competitive interaction leads to phylogenetic clustering. Following this rationale, Mayfield & Levine (2010) broaden the definition of environmental filtering to incorporate not only the abiotic but also the biotic environment, the latter being understood as the interactions that lead to exclusion through differences in competitive ability. This idea is the phylogenetic extension of the sequential deletion rules proposed by Keddy (1992), whereby species that survive the abiotic filter grow and interact, and only those with strong competitive ability persist. We suspect that this nuanced version of the processes leading to phylogenetic clustering might help explain the community structure of soil bacteria, which would be sculpted by two filters. First, the abiotic scenario would pick out the organisms based on their ranges of physiological tolerance to factors such as acidity or salinity (Fierer & Jackson, 2006; Lozupone & Knight, 2007). Second, competition for resources would filter out the weak competitors, *à la* Mayfield & Levine (2010).

Evidence for this latter process has only recently become available. Soils worldwide are dominated by a few lineages, notably the phyla *Proteobacteria* and *Actinobacteria* (Janssen, 2006; Fierer *et al*., 2007). These are powerful competitors in terms of growth response when carbon substrates are supplied to the soil, which is typically carbon-limited (Goldfarb *et al*., 2011). Furthermore, the increased dominance of *Proteobacteria* and *Actinobacteria* in artificially enriched soil microcosms significantly intensifies the phylogenetic clustering of the bacterial community through the competitive exclusion of other phyla (Goldfarb *et al*., 2011). These observations suggest that, analogous to the example in Fig. 1b, the best heterotrophic bacterial competitors have higher growth rates in the presence of carbon substrates, outcompete the rest of the clades and generate phylogenetically clustered patterns.

Here, we perform a literature review to identify general patterns and knowledge gaps on bacterial community phylogenetics with the aims of: (1) assessing with a formal meta-analysis whether soil bacterial communities are phylogenetically clustered; (2) testing if soil bacterial phylogenetic patterns show spatial and/or taxonomic scale dependency; and (3) detecting the abiotic (pH, total organic C and total N) and biotic factors (relative abundance of *Proteobacteria*, *Actinobacteria* and *Acidobacteria*) that underlie the phylogenetic community structure of soil bacteria. We discuss how competitive exclusion of distantly related taxa, based on competitive ability differences, might be a common process driving the community assembly of soil bacteria.

MATERIALS AND METHODS

We performed a formal meta-analysis – a quantitative statistical synthesis of data gathered from multiple studies that address a common question. This procedure allows the combined analysis of data that show heterogeneous variances and are not equally reliable, primarily due to different sampling intensities (Arnqvist & Wooster, 1995; Harrison, 2011). Weighting the studies by sample size (in our case, sequencing depth) is the critical difference between meta-analysis and other quantitative review techniques that can lead to misleading conclusions (Arnqvist & Wooster, 1995). In our study, we followed the typical steps for meta-analyses (Arnqvist & Wooster, 1995; Harrison, 2011), as summarized below.

Selection of studies and measures of effect size

We identified and critically selected the studies, and chose appropriate measures of effect size, that is, statistics providing a standardized measure of the dependent variable across surveys (Arnqvist & Wooster, 1995; Harrison, 2011). We compiled articles that investigated the phylogenetic community structure of soil bacteria, by performing a literature search until September 2012 in Web of Science and Google Scholar, using combinations of keywords ('phylogen*', 'soil', 'bacteria', 'communit*', 'diversity') (see Appendix S1 in Supporting Information for the phases of the information flow through the meta-analysis, depicted as a PRISMA flowchart). We found 44 studies through database searching, and included two additional surveys with unpublished results. After removing duplicates, this resulted in 20 studies, including 330 study cases. We excluded three studies that investigated bacteria in ecosystems other than soil. Finally, we only included those articles that used the net relatedness index (*NRI*) or the nearest taxon index (*NTI*), because these were the most commonly used metrics ($n = 216$ and $n = 211$ respectively; Appendix S2). We excluded one record that used phylogenetic community metrics other than *NRI* or *NTI*, because there were too few case studies for a robust analysis. The final dataset included 231 case studies reported in 16 studies and based on data generated in 43 surveys. Our meta-analysis was constrained to those studies publishing measures of phylogenetic diversity. Calculating these metrics *de novo* would require not only access to the nucleotide sequences, which could be downloaded from public repositories, but also the metadata, which are often difficult to obtain.

Both *NRI* and *NTI* examine whether co-occurring taxa are more (positive values) or less (negative values) closely related than expected by chance. Whereas *NRI* provides information about deep-level relatedness, *NTI* allows a finer-scale phylogenetic inspection (Webb *et al*., 2002). More specifically, *NRI* computes the average of all pairwise phylogenetic distances (mean pairwise distance, MPD) between the taxa in a local community (MPD_{obs}) and compares it to the average of MPD calculated in n randomly constructed communities considering the regional pool of taxa (MPD_{rand}) as follows: $NRI = -(MPD_{obs} - MPD_{rand})/sd\_MPD_{rand}$, where sd_MPD_{rand} is the standard deviation of the MPD_{rand} values (Webb *et al*., 2002). Likewise, *NTI* is a standardized measure of the phylogenetic distance to the nearest taxon (mean nearest-neighbour distance, MNND) for each taxon in a local community ($MNND_{obs}$) compared to that under a null model (as above). The studies

included in our dataset reconstructed phylogenies for the computation of *NRI* and/or *NTI* using neighbour-joining or maximum-likelihood methods in 94% of all study cases, and Bayesian algorithms in the remaining cases. All the selected studies considered the regional pool of taxa as the sum of all taxa detected in the local communities (i.e. in every site included in the study). In 99% of all study cases, *NRI* and/or *NTI* were calculated using null models that generate the same expected phylogenetic distance between any pair of taxa over many randomizations (Kembel, 2009). Particularly, the null communities were constructed by randomly drawing samples either from the local pools ('sample pool'; 58.9% of all cases) or the regional pool ('phylogeny pool'; 30.7%), or by randomly shuffling taxon labels across the tips of the phylogeny ('taxon labels'; 9.5%; Appendix S2).

Defining the meta-analytical model

Once the final set of studies and appropriate measures of effect size (*NRI* or *NTI*) had been selected, we ran Bayesian meta-analyses by fitting generalized linear mixed models (GLMMs) using Markov chain Monte Carlo (MCMC) techniques in the MCMCglmm package for R (Hadfield, 2010; R Development Core Team, 2011). The effect size (*NRI* or *NTI*) was the dependent variable in the model. In order to account for the varying sampling effort across studies, all GLMMs were weighted by n, the number of sequences (Arnqvist & Wooster, 1995), by passing $1/n$ to the '*mev*' function of MCMCglmm (Hadfield & Nakagawa, 2010). In addition, because separate global effect sizes could come from the same publication, we used the publication as a random grouping factor in all calculations. The effect of predictors was estimated by calculating the 95% credible interval of their posterior distributions (Nakagawa & Cuthill, 2007). In all cases, we ran 13,000 MCMC iterations with a burn-in period of 3000 iterations and the default priors. Convergence of the chain was verified by means of the autocorrelation function of the Markov chain.

Testing for confounding effects

Prior to the statistical comparison of effect sizes based on our research questions (see below), we checked whether several methodological variables had any effect on the dependent variable (*NRI* or *NTI*) that describes the phylogenetic community structure of soil bacteria (Verdú & Traveset, 2005). The methodological variability across studies was mainly related to the experimental setup, the target organisms and molecules (all bacteria, a particular taxon or functional guild), and the methods used for community analysis and phylogeny reconstruction (Appendix S2). We found out that none of the methodological variables (i.e. experiment type, marker type, tree completeness, analytical method and phylogeny reconstruction) had a significant effect on *NRI* or *NTI* and thus none of these variables was considered further (Appendix S3).

For the identification of bias against the publication of studies with non-significant effect sizes, we constructed weighted histograms for *NRI* and *NTI* in the weights package for R (Pasek, 2012). In these graphs, the height of the bars reflects the weighted frequency of the studies within each class based on their log-transformed number of sequences (Fig. 2a). The absence of a depression in the region of no effect, i.e. that approaching zero, suggests the absence of publication bias (Rosenberg *et al.*, 2000).

We expected that the sequencing depth would not alter the *NRI* or *NTI* values, because these metrics ensure a within-study standardization of the observed phylogenetic distance based on the sample size through comparison with the expected phylogenetic distance under a null model. This standardization should therefore allow comparisons to be made between studies. We nonetheless confirmed that the effect size was not related to sequencing depth either for *NRI* [MCMCglmm; posterior mean estimate of log(number of sequences) = 0.05, 95% credible interval (−0.27, 0.37), $n = 216$] or *NTI* [−0.02 (−0.46, 0.18), $n = 211$] (Fig. 2b). Similar results were obtained when the weighted histograms and the regressions between *NRI* or *NTI* and the number of sequences were performed exclusively with the set of studies targeting the bacterial domain with phylogenetic markers (Appendix S4) that was used in subsequent analyses (see below).

Statistical comparison of effect sizes

We ran several Bayesian GLMMs as described above in order to address the following questions. First, we evaluated whether soil bacterial communities show a phylogenetic structure that differs from randomness: that is, whether *NRI* or *NTI* depart significantly from zero. Second, we checked if the effect size depends on scale-related variables, namely taxonomic breadth and geographical area (Appendix S2). The latter was the geographical area occupied by all samples considered in the regional pool either as reported in the literature or calculated from the geographical coordinates of the sampling points.

In order to test which abiotic and/or biotic parameters explained the phylogenetic clustering, we used the subset of study cases that exclusively targeted the bacterial domain, and excluded studies focused on a particular taxon or a microbial guild. We performed two separate Bayesian models using (1) only abiotic or (2) abiotic and biotic parameters as independent variables. The abiotic parameters used were pH, total organic C (TOC) and total nitrogen (TN), which are linked to the bacterial community structure (Fierer & Jackson, 2006; Smith *et al.*, 2008; Goberna *et al.*, 2012). The biotic parameters used were the relative abundances of *Proteobacteria*, *Actinobacteria* and *Acidobacteria*, which are typically the three most abundant bacterial phyla in soils (Janssen, 2006; Fierer *et al.*, 2007). The relative abundances of these phyla were used as proxies of their competitive success, because the outcome of competitive interactions is the simultaneous outgrowth and suppression of strong and weak competitors, respectively, as shown experimentally by Goldfarb *et al.* (2011). All abiotic and biotic variables were obtained from the studies included in the meta-analysis or references therein (Appendix S2).

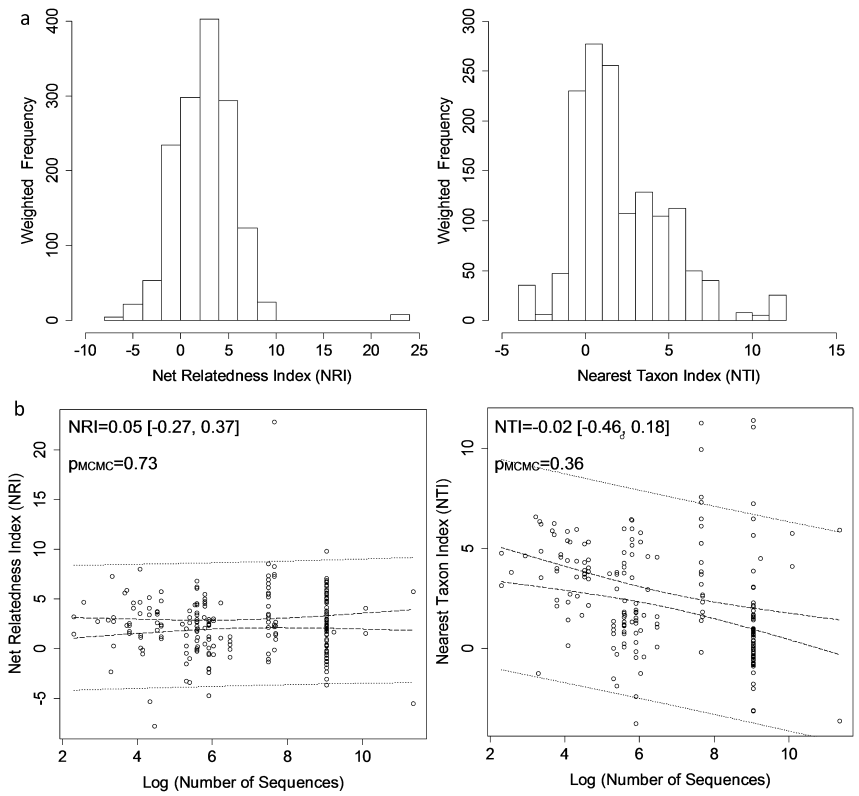

Figure 2 Net relatedness index (*NRI*) and nearest taxon index (*NTI*) of soil bacterial communities worldwide ($n = 216$ and $n = 211$, respectively). *NRI* or *NTI* values indicate phylogenetic clustering (positive values) or overdispersion (negative values). (a) *NRI* and *NTI* weighted histograms by sequencing depth suggest the absence of publication bias in our dataset due to the absence of a depression in the region approaching zero. (b) *NRI* and *NTI* values were not related to the number of sequences in this meta-analysis. Dashed lines show the 95% confidence intervals and the dotted lines indicate the prediction intervals for linear regression.

Other factors known to influence bacterial communities (e.g. soil moisture: Schimel *et al*., 2007) were not considered due to the paucity of data (Appendix S2). The inclusion of soil moisture in the analyses did not alter the conclusions of this study (data not shown).

RESULTS

General patterns

Our dataset contained 231 study cases, covering a wealth of ecosystems including cold and hot deserts, temperate grasslands, tropical rainforests, subalpine and alpine tundra, wetlands, Mediterranean shrublands and arid savannas, among others (Appendix S2). Soils were collected under a broad range of either natural or artificially induced environmental conditions, including extreme temperatures, periodic water saturation, intense radiation, high electrical conductivity, heavy-metal contamination and nitrogen pollution (Appendix S2). Thus, the abiotic soil properties that we explored were highly variable, with pH values ranging from 2.2 to 9.6, total organic C from less than 0.01% to 48% and total N from less than 0.01% to 9% (Appendix S2). We also found a large range of study areas, from 2 m^2 to 5.8×10^{13} m^2. We detected a strong geographical bias towards the Northern Hemisphere (80% of all study cases), with most surveys covering North America and Europe (Appendices S2 & S5).

A total of 56.3% of all study cases investigated soil bacterial communities at the domain level, using universal phylogenetic markers specific to bacteria (mainly the 16S rRNA gene). Only 10.8% of all cases investigated soil bacteria at the phylum level, all of them targeting *Acidobacteria*. The remaining 32.9% of all cases investigated the community structure of nitrogen-cyclers, either using phylogenetic markers (e.g. 16S rRNA gene of ammonia-oxidizing *Betaproteobacteria*) or functional markers of nitrogen fixation (*nifH*) or denitrification (*nirK* and *nirS*). Finally, fewer than half of the study cases were accompanied by a basic dataset with soil variables published in the same or other articles.

The distribution of *NRI* and *NTI* of soil bacterial communities encompassed all overdispersed, random and clustered phylogenetic patterns of community structure (Fig. 2a). Phylogenetic clustering was the general trend, however, because the average *NRI* and *NTI* across studies were significantly greater than zero [*NRI* = 2.29 (1.43, 3.29), $n = 216$; *NTI* = 2.54 (1.63, 3.60), $n = 211$]. Similar figures were obtained when the subset of studies that exclusively targeted the bacterial domain were analysed independently [*NRI* = 1.99 (0.60, 3.52), $n = 115$; *NTI* = 2.56 (1.23, 3.99), $n = 130$].

We found no significant effect of any of the scale-related variables considered on *NRI* or *NTI*, and neither *NRI* [log (area) = 0.037 (–0.053, 0.133), $n = 212$] nor *NTI* [log (area) = 0.085 (–0.009, 0.189), $n = 207$] varied depending on the study area. The same results were found when the studies targeting the bacterial domain were analysed separately [*NRI*: log (area) = –0.03 (–0.22, 0.19), $n = 111$; *NTI*: 0.100 (–0.076, 0.278), $n = 126$]. *NRI* and *NTI* were not affected by the taxonomic breadth either, because there were no significant differences

among taxonomic ranks, as reflected by the overlapping credible intervals of effect sizes (Fig. 3).

Discerning abiotic and biotic factors of phylogenetic clustering

We tested the contributions of several abiotic and biotic predictors to the phylogenetic community structure of soil bacteria. In the multifactorial model that included three abiotic predictive variables, we found a significant and positive effect of total organic C, but not of total N or pH, on both *NRI* and *NTI* (Fig. 4a). In the second model, including all the abiotic and biotic predictors, we found a significant and positive effect of the relative abundance of *Proteobacteria* on both *NRI* and *NTI*, although the relative abundances of *Actinobacteria* and *Acidobacteria*, total organic C, total N and pH showed no significant effect on either dependent variable (Fig. 4b). Proteobacteria were significantly more abundant in soils with higher contents of total organic C [0.307 (0.101, 0.477), $n = 43$]. This relationship mainly responded to the increased abundance of *Alphaproteobacteria* (0.517 [0.191, 0.813]) and *Gammaproteobacteria* [0.371 (0.109, 0.636)]. Finally, we analysed the contribution of the five classes of *Proteobacteria* to the patterns observed. When combining all proteobacterial classes as predictors in a single model, the relative abundance of *Alphaproteobacteria* was the only class that significantly intensified the phylogenetic clustering, measured either as *NRI* [0.165 (0.114, 0.214), $n = 56$] or *NTI* [0.075 (0.026, 0.125), $n = 71$].

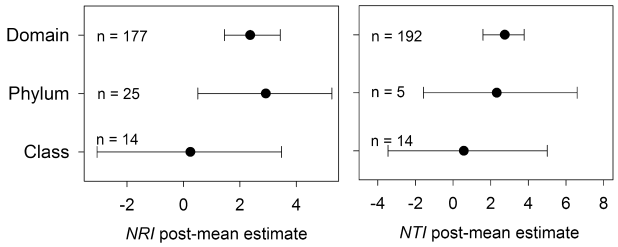

Figure 3 Net relatedness index (*NRI*) and nearest taxon index (*NTI*) of soil bacterial communities at the taxonomic levels studied. Bayesian post-mean estimates and 95% credible intervals for *NRI* and *NTI* values are shown for each taxonomic level. Overlapping intervals across taxonomic levels indicate that *NRI* and *NTI* values did not differ significantly based on the taxonomic breadth. Sample size (n) is included for each group.

DISCUSSION

Soil harbours an enormous bacterial diversity (Curtis *et al.*, 2002), which makes it one of the most diverse ecosystems on Earth. However, our meta-analysis confirmed that soil bacterial communities are significantly less phylodiverse than expected by chance, that is to say, soil bacteria tend to coexist with close relatives. Lozupone & Knight (2007) noticed that, compared to bacterial communities in other natural environments (e.g. water, ice, sediments or mats), soil bacteria show particularly low levels of phylodiversity, contrasting with their enormous species-level diversity. Interpretations of why closely related soil microbes co-occur have often been predicated on the particular abiotic conditions filtering relatives that share traits relevant to environmental tolerance (Horner-Devine & Bohannan, 2006; Bryant *et al.*, 2008; Jones *et al.*, 2009). Some authors have also alluded to methodological drawbacks to explain the prevailing phylogenetic clustering (Horner-Devine & Bohannan, 2006; Bryant *et al.*, 2008). We discuss these methodological aspects below, and argue that filtering is not only mediated by abiotic factors but also by biotic interactions (*sensu* Mayfield & Levine, 2010).

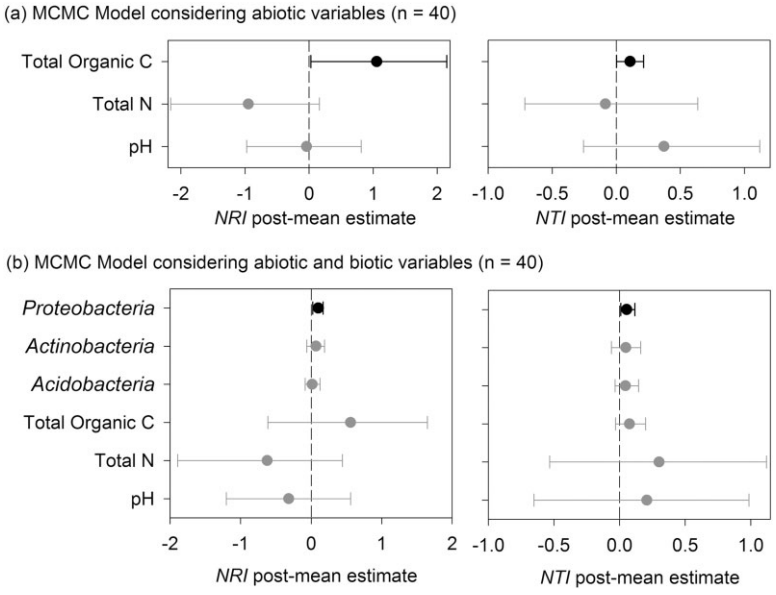

Figure 4 Phylogenetic community structure of soil bacteria explained by (a) abiotic or (b) abiotic and biotic factors. Bayesian post-mean estimates and 95% credible intervals for net relatedness index (*NRI*) and nearest taxon index (*NTI*) values are shown. (a) In a single model including all abiotic variables as predictors, total organic C was the only factor significantly explaining *NRI* and *NTI* (black intervals), whereas the remaining abiotic factors did not have a significant effect on *NRI* or *NTI* (grey intervals). (b) In a single model using all abiotic and biotic variables as predictors, proteobacterial relative abundance was the only factor significantly explaining *NRI* and *NTI* values. Note that these analyses were performed only with the set of study cases considering the whole bacterial domain for which abiotic and biotic variables were available. Total organic C and total N were log-transformed for models using *NRI* as dependent variable.

Methodological artefacts and phylogenetic clustering

One methodological issue that has been raised suggests that certain phylogenetic markers that are used to identify the community members might produce phylogenetic but not functional clustering. That is to say, co-occurring organisms are phylogenetically closely related but they are not functionally similar. This might occur if close relatives segregate their niches due to horizontal gene transfer (HGT), a process that would not be reflected in their evolutionary distance, defined as 16S rRNA gene sequence identity (Lozupone & Knight, 2007). This envisages HGT as being so rampant that phylogeny (reconstructed based on the 16S rRNA gene) does not reflect function. Although the contribution of HGT to bacterial genome evolution is still under debate, the fixation of transferred genes is not substantial between phylogenetically distant organisms (Kurland *et al*., 2003; Choi & Kim, 2007). This observation is consistent with the congruence between trees based on the 16S rRNA gene and those based on whole-genome gene-contents (Snel *et al*., 1999), and consequently with the significant relationship between bacterial phylogeny and function (Barberán *et al*., 2014). Another misleading use of phylogenetic clustering as a proxy of functional clustering might arise from the fact that bacteria identified through their rRNA genes can be active in different temporal windows. In other words, the rRNA gene (DNA) reflects not only active community members but also dormant individuals. However, DeAngelis & Firestone (2012) demonstrated that phylogenetic community patterns based on RNA, i.e. culling only the active microbes, can be significantly more clustered than those based on DNA. Phylogenetic clustering has been shown to be also dominant when functional rather than phylogenetic markers are used for bacterial community analysis (Jones & Hallin, 2010; Hamilton *et al*., 2011). Indeed, in our meta-analysis we did not find significant differences in *NRI* or *NTI* based on the use of phylogenetic *versus* functional markers. Our dataset included two types of functional markers, namely genes coding for enzymes involved in nitrogen fixation and denitrification. The evolution of these genes does not differ significantly from that expected under a Brownian model (Martiny *et al*., 2013). This indicates that the phylogenetic conservatism of the functional markers that we analysed does not differ significantly from that of the conserved taxonomic marker (16S rRNA gene). This agrees with the significant relationship found between 16S rRNA distance and functional distance based on protein families' domains (Barberán *et al*., 2014). Kembel *et al*. (2011) also reported concordance between *NRI* values calculated based on the 16S rRNA gene and on multigene metagenomic phylogenies. Overall, phylogenetically clustered patterns prevail regardless of the molecular marker used to investigate soil bacteria, indicating that phylogenetic clustering reflects a functional process at the community level.

Another methodological constraint has been associated with the notion that phylogenetic community patterns are scale-sensitive (Horner-Devine & Bohannan, 2006; Bryant *et al*., 2008). In plant communities, it has been shown that the larger the working area the stronger the phylogenetic clustering (Swenson *et al*., 2006). The logic behind this tendency is that as the regional taxon pool is scaled out, it is more likely that the randomly assembled communities will include higher-level clades that are absent in the local community. This process can be remarkable in organisms for which the number of taxa increases rapidly with the area, i.e. those with a steep slope of the taxa–area plot, as is the case for plants (Horner-Devine *et al*., 2004). We found no scale-dependency for the phylogenetic community structure of soil bacteria, however, either considering geographical or taxonomic scale-related variables. This agrees with the observation that bacteria have the lowest rate of change in their taxa–area relationship of any group of organisms (Horner-Devine *et al*., 2004). Taking the figures provided by these authors, an increase in area similar to that in our dataset ($2–10^{14}$ m^2) would augment the number of taxa 85 times more for plants than for bacteria (defined as operational taxonomic units at 97% sequence identity). Arguably, the taxa–area relationships calculated for bacteria might have been underestimated because infrequent taxa are undersampled due to the difficulties inherent in studying the vast diversity of soil microbes (Woodcock *et al*., 2006). The lack of correlation between *NRI* and *NTI* with sequencing depth in our study suggests that adding infrequent taxa would not alter the phylogenetic community structure.

Phylogenetic clustering driven by abiotic and biotic filtering

Abiotic factors, such as acidity, salinity or moisture are the primary determinants of bacterial community structure (Fierer & Jackson, 2006; Lozupone & Knight, 2007; Schimel *et al*., 2007). Abiotic filtering leads to the overrepresentation of certain clades given their tolerance to particular environmental conditions. As an illustration, *Acidobacteria* were more abundant in our dataset at decreasing pH values [multifactorial MCMCglmm; estimate = −0.32 (−0.71, −0.02); $n = 40$]. Indeed, Jones *et al*. (2009) showed that acidobacterial phylogenetic clustering increases as pH deviates from neutrality at the continental scale.

Our meta-analysis suggests, however, that biotic filtering can be also relevant in shaping bacterial communities. We found that the overrepresentation of the efficiently carbon-using proteobacteria can partly explain the widespread coexistence of phylogenetically related soil bacteria. Soil productivity intensifies such a pattern at a worldwide scale because proteobacteria are increasingly favoured at high resource availability. These observations, supported by an independent line of experimental evidence (Goldfarb *et al*., 2011), are consistent with phylogenetic clustering arising via the competitive exclusion of deeply branching clades with low competitive abilities (Mayfield & Levine, 2010).

The relative abundance of proteobacteria, which we used as a surrogate of their competitive success, was the best predictor of the phylogenetic clustering among a set of abiotic and biotic factors known to influence the community assembly of soil bacteria. Proteobacterial abundance, particularly that of

Alphaproteobacteria and *Gammaproteobacteria*, was significantly associated with the levels of total organic C. Proteobacteria have been acknowledged before to be reflective of the natural soil's fertility and respond strongly to artificial C amendments (Fierer *et al.*, 2007; Philippot *et al.*, 2010; Goldfarb *et al.*, 2011; Ganz *et al.*, 2012). This has led to their consideration as copiotrophs, with high maximum growth rates and hence competitive advantage at high resource availability (Fierer *et al.*, 2007). Importantly, the bacterial growth response to C substrates is phylogenetically conserved across deep evolutionary clades (Goldfarb *et al.*, 2011). In a striking experiment, proteobacteria were found to be the strongest competitors in terms of growth response when labile C substrates, similar to root exudates, were supplied to the soil (Goldfarb *et al.*, 2011). More specifically, over 300 taxa grew in response to the addition of either sucrose or glycine, of which 78% were *Proteobacteria* and 21% *Actinobacteria*. Furthermore, the mentioned single-C-source amendments suppressed bacterial taxa belonging to over 20 phyla other than *Proteobacteria* or *Actinobacteria* (Goldfarb *et al.*, 2011), indicating that C generates asymmetric competition among whole bacterial clades. Such a competitive exclusion of entire deeply branching clades from the bacterial tree drastically reduces the phylogenetic diversity of the community, and ultimately generates phylogenetic clustering.

Addressing future questions

Our meta-analysis has uncovered general trends that may explain the phylogenetic structure of bacterial communities but, at the same time, it has detected gaps of knowledge that future research should address. First, a thorough analysis of the spatial and taxonomic scales is needed, including smaller sampling scales (below the square metre) and finer taxonomic ranks (below the class level), especially given that it seems to be a tendency for lower taxonomic levels to be less clustered. Second, it should be investigated whether methodological differences in OTU delimitation (e.g. Koeppel & Wu, 2013) influence the estimates of phylogenetic community structure by better capturing the ecologically relevant units. Third, it should be noticed that the relative abundances of bacterial phyla are calculated from the number of 16S rRNA gene copies, which can vary from one up to 15 copies (Pei *et al.*, 2010). Recent mathematical simulations have shown that using gene-copy numbers instead of individual numbers underestimates the relative abundance of the most abundant taxa (Kembel *et al.*, 2012). On average, proteobacteria have a comparatively high number of copies of the 16S rRNA gene (Pei *et al.*, 2010). Hence, we expect that correcting our dataset for the 16S rRNA gene copy numbers would further strengthen the phylogenetic clustering and increase the effect of proteobacterial abundance. The interspecies gene copy number variation should be accounted for in future works inferring phylogenetic community structure (Kembel *et al.*, 2012). Future studies should also address the complexity of ecological processes, such as predation, syntrophy or competitive exclusion by limiting similarity, that simultaneously filter all functional traits involved in the survival and adaptation of the multitude of taxa that shape the biological communities (Mayfield *et al.*, 2009). Finally, it should be investigated whether local diversification influences the phylogenetic community structure of soil bacteria, because this process may produce phylogenetic clustering (Pausas & Verdú, 2010). We suspect that if the effect of *in situ* diversification were relevant, then soil bacterial community assembly would be more closely related to geographical distance (and less to environmental distance) than shown by current observations (Horner-Devine *et al.*, 2004; Fierer & Jackson, 2006; Martiny *et al.*, 2006). Bacteria are ideal organisms to perform long-term experimental evolution, which can now be surveyed integrating functional and genomic information, thus providing a promising scenario to discern the relative magnitude of ecological and evolutionary forces that underlie community assembly processes.

ACKNOWLEDGEMENTS

We are very grateful to E. S. Boyd, E. Costello, C. W. Chong, K. M. DeAngelis and H. H. Ganz for kindly providing us with further information on their studies. We also thank Pedro Pérez Cutillas for helping us with Appendix S5. Financial support was provided by the EU Marie Curie Programme (FP7-PEOPLE-2009-RG-248155) and the Spanish Ministry of Science and Innovation (R+D Project CGL2011-29585-C02-01 and CSIC JAE-Doc Programme cofunded by the European Social Fund). Thoughtful comments by anonymous referees are gratefully acknowledged.

REFERENCES

Arnqvist, G. & Wooster, D. (1995) Meta-analysis: synthesizing research findings in ecology and evolution. *Trends in Ecology and Evolution*, **10**, 236–240.

Baas Becking, L.G.M. (1934) *Geobiologie of inleiding tot de milieukunde*. W.P. van Stockum & Zoon, The Hague.

Barberán, A., Ramirez, K.S., Leff, J.W., Bradford, M.A., Wall, D.H. & Fierer, N. (2014) Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecology Letters*, **17**, 794–802.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11505–11511.

Chesson, P. (2000) Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, **31**, 343–366.

Choi, I.-G. & Kim, S.-H. (2007) Global extent of horizontal gene transfer. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 4489–4494.

Chong, C.W., Pearce, D.A., Convey, P., Yew, W.C. & Tan, I.K.P. (2012) Patterns in the distribution of soil bacterial 16S rRNA gene sequences from different regions of Antarctica. *Geoderma*, **181-182**, 45–55.

Costello, E.K., Halloy, S.R.P., Reed, S.C., Sowell, P. & Schmidt, S.K. (2009) Fumarole-supported islands of biodiversity

within a hyperarid, high-elevation landscape on Socompa Volcano, Puna de Atacama, Andes. *Applied and Environmental Microbiology*, **75**, 735–747.

Curtis, T.P., Sloan, W.T. & Scanell, J.W. (2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 10494–10499.

DeAngelis, K.M. & Firestone, M.K. (2012) Phylogenetic clustering of soil microbial communities by 16S rRNA but not 16S rRNA genes. *Applied and Environmental Microbiology*, **78**, 2459–2461.

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

Fierer, N., Bradford, M.A. & Jackson, R.B. (2007) Toward and ecological classification of soil bacteria. *Ecology*, **88**, 1354–1364.

Ganz, H.H., Karaoz, U., Getz, W.M., Versfeld, W. & Brodie, E.L. (2012) Diversity and structure of soil bacterial communities associated with vultures in an African savanna. *Ecosphere*, **3**, 1–18.

Goberna, M., García, C., Insam, H., Hernández, M.T. & Verdú, M. (2012) Burning fire-prone Mediterranean shrublands: immediate changes in soil microbial community structure and ecosystem functions. *Microbial Ecology*, **64**, 242–255.

Goldfarb, K.C., Karaoz, U., Hanson, C.A., Santee, C.A., Bradford, M.A., Treseder, K.K., Wallenstein, M.D. & Brodie, E.L. (2011) Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Frontiers in Microbiology*, **2**, 94.

Hadfield, J.D. (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.

Hadfield, J.D. & Nakagawa, S. (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, **23**, 494–508.

Hamilton, T.L., Boyd, E.S. & Peters, J.W. (2011) Environmental constraints underpin the distribution and phylogenetic diversity of *nifH* in the Yellowstone geothermal complex. *Microbial Ecology*, **61**, 860–870.

Harrison, F. (2011) Getting started with meta-analysis. *Methods in Ecology and Evolution*, **2**, 1–10.

van der Heijden, M.G.A., Bardgett, R.D. & van Straalen, N.M. (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, **11**, 296–310.

Horner-Devine, M.C. & Bohannan, B.J.M. (2006) Phylogenetic clustering and overdispersion in bacterial communities. *Ecology*, **87**, S100–S108.

Horner-Devine, M.C., Lage, M., Hughes, J.B. & Bohannan, B.J.M. (2004) A taxa–area relationship for bacteria. *Nature*, **432**, 750–753.

Janssen, P.H. (2006) Identifying dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology*, **72**, 1719–1728.

Jones, C.M. & Hallin, S. (2010) Ecological and evolutionary factors underlying global and local assembly of denitrifier communities. *ISME Journal*, **4**, 633–641.

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R. & Fierer, N. (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME Journal*, **3**, 442–453.

Keddy, P.A. (1992) Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science*, **3**, 157–164.

Kembel, S.W. (2009) Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecology Letters*, **12**, 949–960.

Kembel, S.W., Eisen, J.A., Pollard, K.S. & Green, J.L. (2011) The phylogenetic diversity of metagenomes. *PLoS ONE*, **6**, e23214.

Kembel, S.W., Wu, M., Eisen, J.A. & Green, J.L. (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology*, **8**, e1002743.

Koeppel, A.F. & Wu, M. (2013) Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Research*, **41**, 5175–5188.

Kurland, C.G., Canback, B. & Berg, O.G. (2003) Horizontal gene transfer: a critical review. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9658–9662.

Lozupone, C.A. & Knight, R. (2007) Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 11436–11440.

Martiny, A.C., Treseder, K. & Pusch, G. (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME Journal*, **7**, 830–838.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Øvreås, L., Reysenbach, A.-L., Smith, V.H. & Staley, J.T. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, **4**, 102–112.

Mayfield, M.M. & Levine, J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of the communities. *Ecology Letters*, **13**, 1085–1093.

Mayfield, M.M., Boni, M.F. & Ackerly, D.D. (2009) Traits, habitats, and clades: identifying traits of potential impact to environmental filtering. *The American Naturalist*, **174**, E1–E22.

Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591–605.

Pasek, J. (2012) *weights: weighting and weighted statistics*. R package version 0.75. Available at: http://cran.r-project.org/web/packages/weights/index.html

Pausas, J.G. & Verdú, M. (2010) The jungle of methods for evaluating phenotypic and phylogenetic structure of communities. *BioScience*, **60**, 614–625.

Pei, A.Y., Oberdorf, W.E., Nossa, C.W., Agarwal, A., Chokshi, P., Gerz, E.A., Jin, Z., Lee, P., Yang, L.-Y., Poles, M., Brown, S.M., Sotero, S., DeSantis, T., Brodie, E., Nelson, K. & Pei, Z.-H. (2010) Diversity of 16S rRNA genes within individual prokaryotic genomes. *Applied and Environmental Microbiology*, **76**, 3886–3897.

Philippot, L., Andersson, S.G.E., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B. & Hallin, S. (2010) The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology*, **8**, 523–529.

R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rosenberg, M.S., Adams, D.C. & Gurevitch, J. (2000) *metawin: statistical software for meta-analysis*, Version 2.0. Sinauer Associates, Sunderland, MA.

Schimel, J., Balser, T.C. & Wallenstein, M. (2007) Microbial stress-response physiology and its implications for ecosystem function. *Ecology*, **88**, 1386–1394.

Smith, N.R., Kishchuk, B.E. & Mohn, W.W. (2008) Effects of wildfire and harvest disturbances on forest soil bacterial communities. *Applied and Environmental Microbiology*, **74**, 216–224.

Snel, B., Bork, P. & Huynen, M.A. (1999) Genome phylogeny based on gene content. *Nature Genetics*, **21**, 108–110.

Swenson, N.G., Enquist, B.J., Pither, J., Thompson, J. & Zimmerman, J.K. (2006) The problem and promise of scale dependency in community phylogenetics. *Ecology*, **87**, 2418–2424.

Verdú, M. & Traveset, A. (2005) Early emergence enhances plant fitness: a phylogenetically controlled meta-analysis. *Ecology*, **86**, 1385–1394.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

Woodcock, S., Curtis, T.P., Head, I.M., Lunn, M. & Sloan, W.T. (2006) Taxa–area relationship for microbes: the unsampled and the unseen. *Ecology Letters*, **9**, 805–812.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 PRISMA diagram.
Appendix S2 Dataset used to perform the meta-analysis.
Appendix S3 Testing for confounding effects.
Appendix S4 Results with the set of studies targeting the bacterial domain.
Appendix S5 World map with sites included in the meta-analysis.

BIOSKETCHES

Marta Goberna is a microbial ecologist interested in the relationship between soil microbial community structure and the performance of ecosystems.

Carlos García is a soil and organic waste biochemist interested in the biogeochemical cycles.

Miguel Verdú is an evolutionary ecologist interested in the ecological interactions assembling the structure of communities.

Editor: Ian Wright