ENVIRONMENTAL STUDIES

Biogeography of soil bacteria and archaea across France

Battle Karimi[1]*, Sébastien Terrat[1]*, Samuel Dequiedt[1], Nicolas P. A. Saby[2], Walid Horrigue[1], Mélanie Lelièvre[3], Virginie Nowak[1], Claudy Jolivet[2], Dominique Arrouays[2], Patrick Wincker[4], Corinne Cruaud[4], Antonio Bispo[2], Pierre-Alain Maron[1], Nicolas Chemidlin Prévost-Bouré[1], Lionel Ranjard[1†]

Over the last two decades, a considerable effort has been made to decipher the biogeography of soil microbial communities as a whole, from small to broad scales. In contrast, few studies have focused on the taxonomic groups constituting these communities; thus, our knowledge of their ecological attributes and the drivers determining their composition and distribution is limited. We applied a pyrosequencing approach targeting 16*S* ribosomal RNA (rRNA) genes in soil DNA to a set of 2173 soil samples from France to reach a comprehensive understanding of the spatial distribution of bacteria and archaea and to identify the ecological processes and environmental drivers involved. Taxonomic assignment of the soil 16*S* rRNA sequences indicated the presence of 32 bacterial phyla or subphyla and 3 archaeal phyla. Twenty of these 35 phyla were cosmopolitan and abundant, with heterogeneous spatial distributions structured in patches ranging from a 43- to 260-km radius. The hierarchy of the main environmental drivers of phyla distribution was soil pH > land management > soil texture > soil nutrients > climate. At a lower taxonomic level, 47 dominant genera belonging to 12 phyla aggregated 62.1% of the sequences. We also showed that the phylum-level distribution can be determined largely by the distribution of the dominant genus or, alternatively, reflect the combined distribution of all of the phylum members. Together, our study demonstrated that soil bacteria and archaea present highly diverse biogeographical patterns on a nationwide scale and that studies based on intensive and systematic sampling on a wide spatial scale provide a promising contribution for elucidating soil biodiversity determinism.

INTRODUCTION

Soil is the most complex environment on Earth and hosts huge bacterial abundance and diversity with about 10^9 to 10^{10} cells and 10^5 to 10^6 unique "taxonomic groups" in a single gram of soil (*1*). Numerous studies during the last decade have demonstrated the role of soil bacterial diversity (that is, richness, evenness, and community structure) in soil functions, such as nutrient cycling, pathogen management, degradation of pollutants, soil structure improvement, and stability of other ecosystem services to environmental changes [see (*2*) for a review]. Given the key role of soil microorganisms in the regulation of soil ecosystem functions, the environmental factors driving soil bacterial diversity need to be understood (*3*). A large body of data collected at various spatial scales suggests that the diversity and assemblages of soil bacterial and archaeal communities are mainly determined by soil properties (for example, pH, carbon content, texture), land management, climate, and plant cover (*1*, *3*–*6*). However, the processes and drivers influencing the abundance of individual bacterial and archaeal taxa are not clearly understood (*7*). Some microbial taxa are cosmopolitan and can be found in a large range of environmental conditions, whereas other taxa are more specialized and depend on a far more restricted range of environmental conditions. Unfortunately, the environmental drivers that shape each microbial taxon remain unidentified, which hampers our ability to predict their likely variations in a changing environment (*3*).

In this context, a pioneering study involving comparison of 71 soil samples from a wide range of North American soil ecosystems demonstrated that bacterial taxa can be classified into ecologically meaningful categories based on copiotrophic and oligotrophic attributes (*8*). Other more recent studies conducted on a broad spatial scale were focused on specific taxa, but less intensively than the research reported here. For Actinobacteria, one of the dominant bacteria in soils, abundance was shown to be primarily driven by latitude and secondarily by pH, whereas climatic factors did not have any influence (*9*). Other dominant phyla, such as Proteobacteria, Acidobacteria, Planctomycetes, Bacteroidetes, and Firmicutes, showed different sensitivities to pH, C/N ratio, and phosphorous content (*10*). Moreover, Verrucomicrobia and Gemmatimonadetes, two phyla less abundant in soils, were sensitive to land management and soil moisture, respectively (*11*). The main ecological filters for archaeal populations in soils were identified as the C/N ratio and organic carbon content (*12*). Together, these different studies showed that each phylum is associated with specific drivers, which emphasizes the need for precise investigation of the processes and drivers involved in their regulation. It is now acknowledged that certain high-level bacterial taxa at high taxonomic levels (for example, phylum or subphyla) can display shared ecological characteristics since they respond predictably to environmental variables and carry important biological functions (*13*). However, most studies have been limited to a handful of taxa (usually a single phylum) and based only on a few tens of soil samples (*14*), thus limiting the generality of those abiotic parameters as driving microbial populations. In addition, soil samples were generally collected in a more or less restricted area (*10*) or directed to a precise environmental issue, for example, along a steep precipitation gradient

[1] Agroécologie, AgroSup Dijon, Institut National de la Recherche Agronomique (INRA), Université Bourgogne Franche-Comté, F-21000 Dijon, France. [2] INRA Orléans, US 1106, Unité INFOSOL, Orléans, France. [3] Agroécologie–Plateforme GenoSol, BP 86510, F-21000 Dijon, France. [4] Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Biologie François Jacob, Genoscope, 2, Rue Gaston Crémieux, CP5706, 91057 Evry cedex, France.
*These authors contributed equally to this work.
†Corresponding author. Email: lionel.ranjard@inra.fr

(*15*). To tackle these limitations, it is crucial to better integrate all the dominant and minor taxa constituting the community, as well as a wide range of environmental parameters such as soil types, land management, climate, and geography. To reach this goal, one of the most promising strategies is to apply a strong sampling effort in terms of number of soils and microbial taxonomic analyses of the indigenous communities on a broad spatial scale.

We therefore conducted an investigation using a national soil survey, the French Soil Quality Monitoring Network [Réseau de Mesures de la Qualité des Sols (RMQS)], that covers the huge environmental diversity across the whole of continental France (2173 soils, area covered ≈ 5.3 10^5 km^2; Fig. 1) (*16*). The variations in bacterial and archaeal phyla in all these soils were assessed by a pyrosequencing approach targeting 16*S* ribosomal RNA (rRNA) genes. A geostatistical approach was then used to map and compare the spatial distribution of each identified phylum across France. A variance partitioning analysis was also applied to identify and rank the ecological processes and environmental drivers involved in bacterial and archaeal phyla distribution. This approach enabled us to decipher the spatial distribution and environmental drivers of dominant and minor soil bacteria and archaea. Finally, to disentangle the phyla distribution in light of the ecology of groups identified at a lower taxonomic rank, we extended our analysis to the dominant genera detected and identified in French soils.

RESULTS

Ubiquity and dominance of bacteria and archaea

Among the 37 phyla (32 bacteria and 5 archaea) and 2028 genera in the SILVA database, the taxonomic assignment of the data set indicated the presence of 32 bacterial taxa (encompassing 27 phyla and 5 subphyla) and 3 archaeal phyla in sampled soils (Fig. 2) and 1355 genera detected in at least two soils. Comparison of the average relative abundance and ubiquity of the 35 phyla recorded led to the definition of four groups: major, medium, minor, and rare phyla (Fig. 2 and fig. S1). The 15 phyla in the first two groups were recorded in all soils, with an average relative abundance ranging from 14 to 0.5% of sequences per sample. Bacteroidetes, Alphaproteobacteria, Actinobacteria, Planctomycetes, Firmicutes, Gammaproteobacteria, Acidobacteria, and Betaproteobacteria are the most dominant and cosmopolitan phyla. In contrast, the rare phyla (for example, Chlamydiae, Thermotogae, and Fusobacteria) were detected in fewer than 50% of the soils, with less than 0.01% on average per soil sample (Fig. 2).

Among the 1355 genera referenced, only 47 exhibited an average relative sequence abundance greater than 0.5%. All these genera were considered dominant since, on average, they represented more than 62% of the cumulated sequences per sample. These 47 dominant genera belonged to 12 phyla (Fig. 3 and table S4), previously described and classified as the major and medium phyla. The most abundant genus

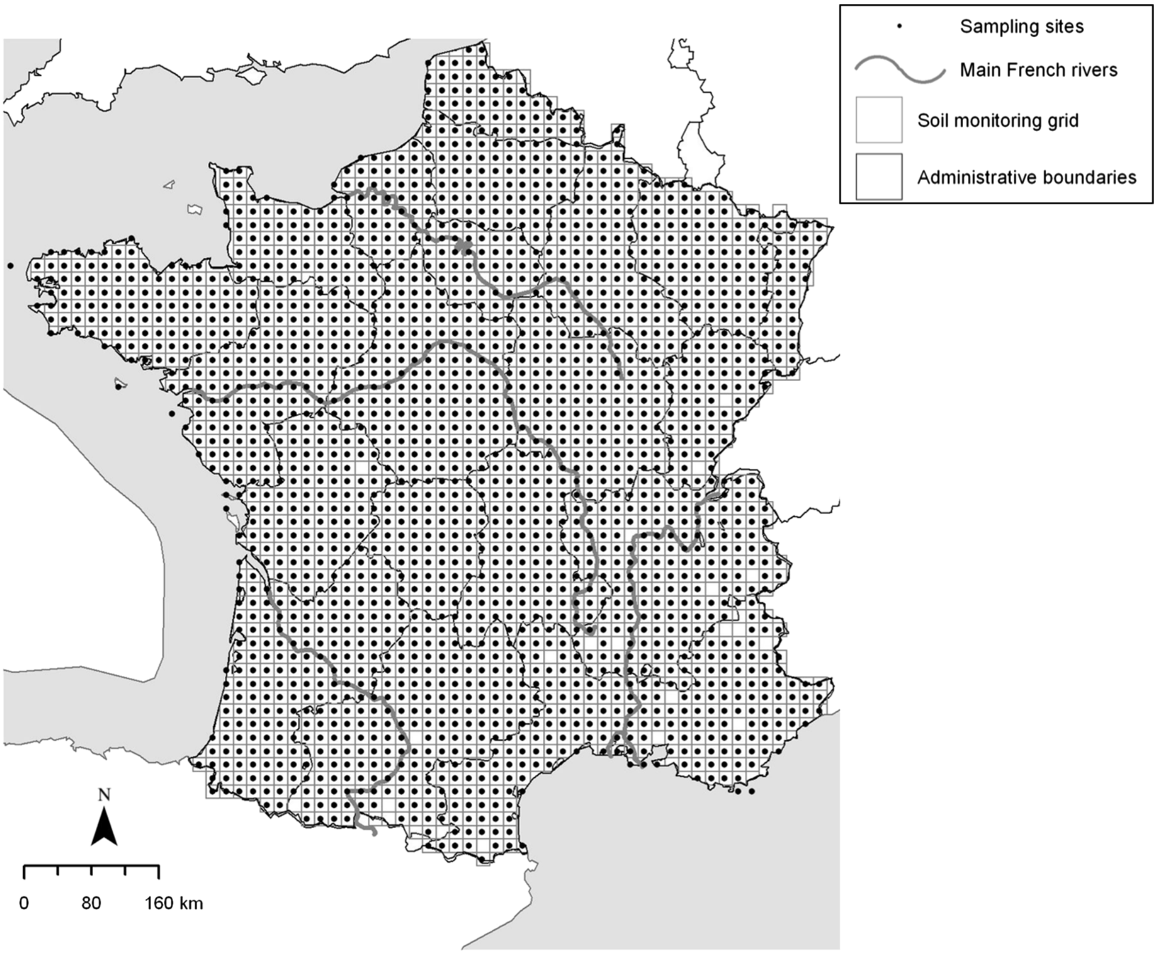

Fig. 1. Sampling design. Map of France and the systematic sampling grid (16 × 16 km) of the French Soil Quality Monitoring Network (RMQS) (*16*).

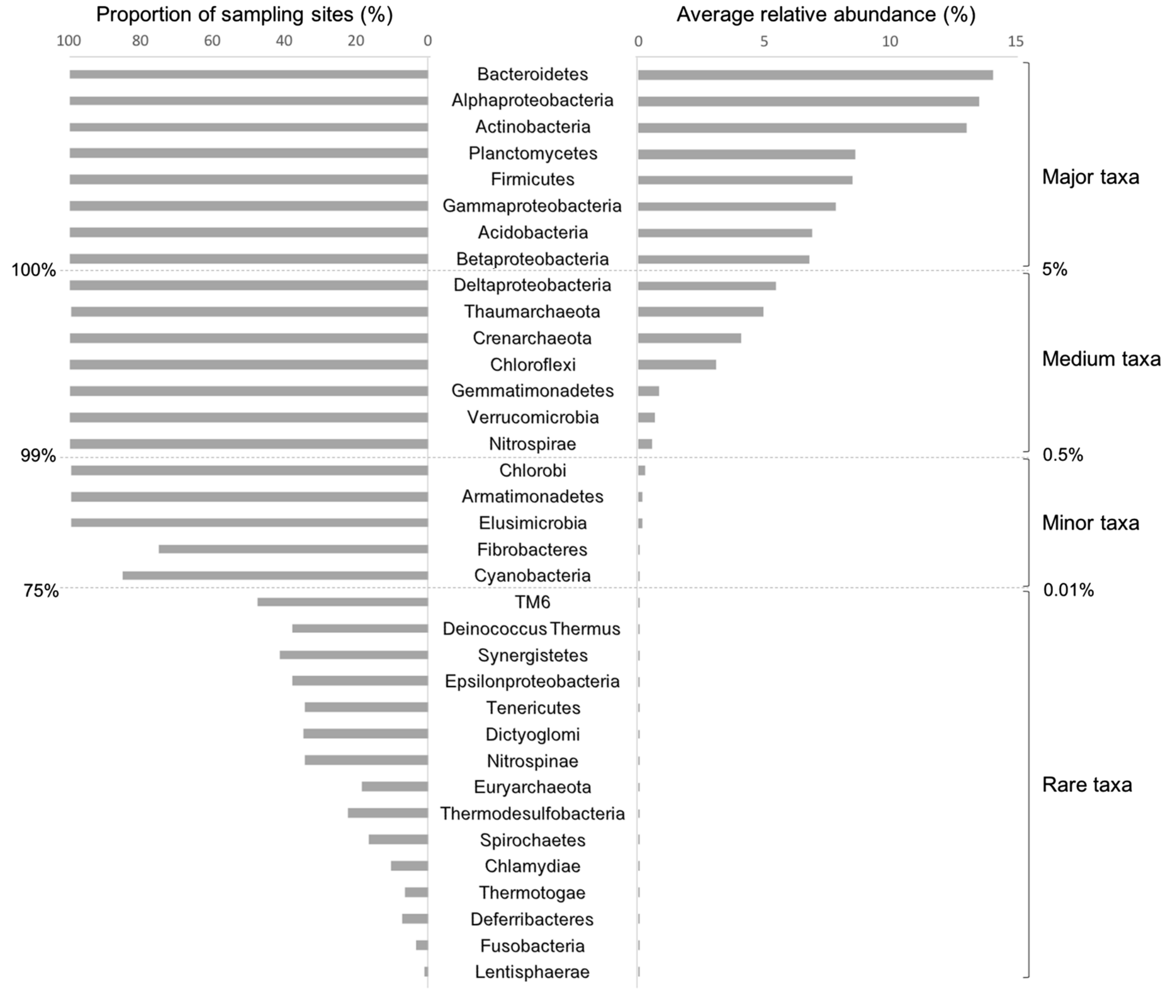

Fig. 2. Representativeness of bacterial and archaeal phyla in French soils. Left: The proportion of sampling sites where phyla were present. Right: The average relative abundance of the phyla. The four groups were determined by ascendant hierarchical clustering (fig. S1). The statistical differences in the phyla distributions are indicated in table S1.

was *Holophaga*, belonging to Acidobacteria, with an average relative abundance of 6.29% (table S4), and the 47th genus was *Desulfobulbus*, belonging to Deltaproteobacteria, with an average relative abundance of 0.52%. On average, these 47 genera occurred in 98.2% of the samples, and the least cosmopolitan in the sampling area (*Ktedonobacter*, member of Chloroflexi) was found in 79.6% of the samples.

Mapping of bacteria and archaea

Maps of the 20 most representative phyla, identified in Fig. 2, were drawn by applying a geostatistical approach based on their relative abundance (Fig. 4, fig. S2, and table S2). In the linear models used for interpolation, R^2 ranged from 0.04 to 0.29, revealing spatial structuring depending on the phylum. Each phylum exhibited a heterogeneous and spatially structured distribution, and three types of distribution patterns were distinguished for the different phyla, based on the size of the geographical patches that ranged from 43.3- to 260.2-km radius. Chloroflexi, Fibrobacteres, and Cyanobacteria exhibited the spottiest distribution with patches around 50-km radius, whereas Actinobacteria, Firmicutes, Gammaproteobacteria, Betaproteobacteria, Nitrospirae, Chlorobi, and Elusimicrobia exhibited the patchiest distribution with patches larger than 200 km radius. The other 10 phyla gave intermediate-sized patches of about 100- to 150-km radius (Fig. 4 and fig. S2).

The genera distributions were more spatially structured than the phyla distributions (Fig. 5, fig. S3, and table S4). The spatial models of 60% of the phyla showed an R^2 greater than 10%, compared with 80% of the genera. Three of the 47 genera (*Gaiella*, *Phycisphaera*, and *Thermofilum*) were not spatially structured, and the other 44 had significant spatial distributions with geographical patches ranging from 25- to 314.5-km radius. For seven phyla, the spatial distribution of the most abundant genus was sufficient to explain the spatial distribution of the phylum, for example, *Holophaga* for Acidobacteria, *Terrimonas* for Bacteroidetes, *Bacillus* for Firmicutes, and *Sulfobococcus* for Crenarchaeota (Fig. 5A). However, as illustrated by the Actinobacteria and Deltaproteobacteria phyla (Fig. 5B), the spatial distribution of the phylum was not systematically driven by the major genus but instead by the cumulative distributions of all the genera.

Relationship between environmental parameters and bacteria and archaea distributions

The total explained variance in phyla distribution ranged from 17% (Cyanobacteria) to 60% (Alphaproteobacteria and Bacteroidetes)

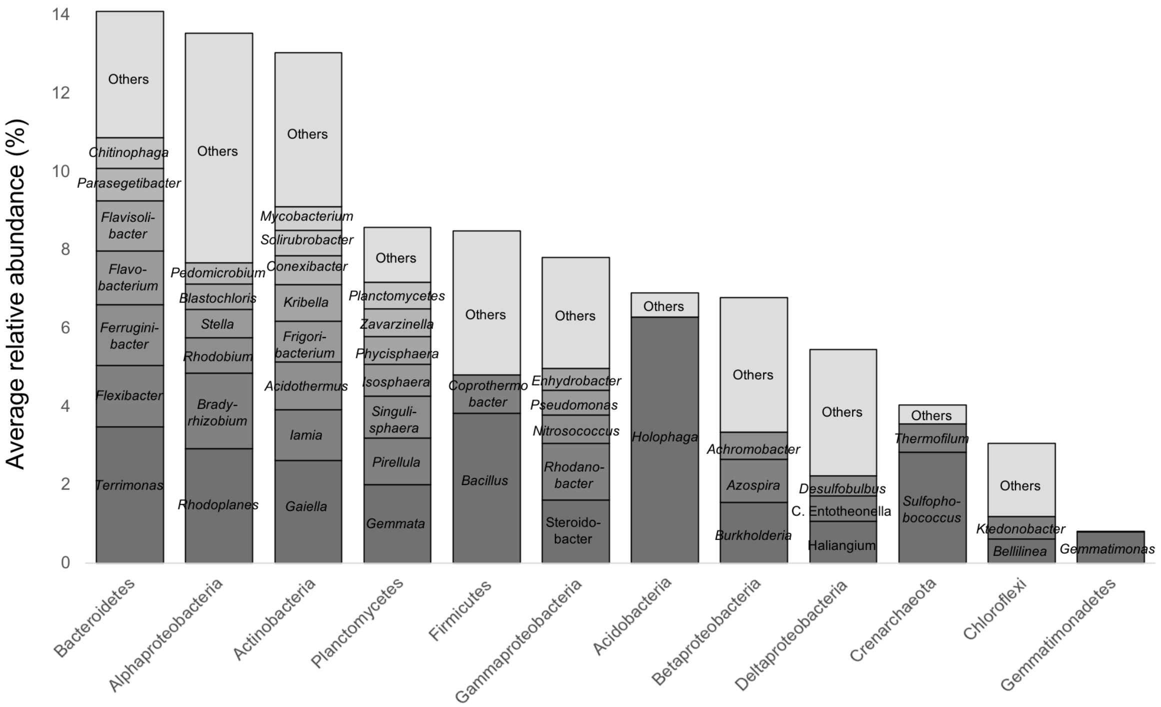

Fig. 3. Relative abundance of the 47 main genera. The main genera presented more than 0.5% of sequences on average and occurred in more than 75% of sites. The genera were classified across the bacterial and archaeal phyla. "Others" corresponds to the sum of all genera within each phylum representing fewer than 0.5% of sequences.

(Fig. 6A and table S3). The residual variance (from 40 to 83%) indicates that other parameters, not taken into account in our study, may also contribute to the distribution of bacterial and archaeal phyla in these samples. Total variance was partitioned between five types of explanatory sets: four sets of environmental parameters (soil properties, land management, climate, and interactions between environmental properties) and one set of spatial descriptors [spatial location, elevation, and principal coordinates of neighbor matrix (PCNM) variables]. For 16 phyla, the influence of selection by soil type, land management, and their interactions was stronger than the influence of spatial descriptors (Fig. 5A). These selection processes influenced both the major and medium phyla, for example, Bacteroidetes and Alphaproteobacteria, and the minor phyla, for example, Elusimicrobia and Fibrobacteres. Soil properties were a main factor in the selection process for 9 of these 16 phyla and explained 7.8 to 27.9% of the total variance. Land management was the main factor for two major and four minor phyla (from 8.0% of the explained variance for Gemmatimonadetes to 12.7% for Alphaproteobacteria; Fig. 6A). Finally, climate represented the weakest selective pressure (less than 3.2% of the explained variance) and concerned only seven phyla (five minor and two abundant). In addition, interactions between environmental parameters explained between 0 and 27.8% of total variance according to the phylum. On the other hand, spatial descriptors mainly explained the variations (from 6.2 to 17.0%) of four phyla, two of which were dominants (that is, Betaproteobacteria and Firmicutes) and two were minors (that is, Chloroflexi and Cyanobacteria) (Fig. 6A).

When the relative contribution of each environmental parameter was ranked according to the respective amounts of explained variance, the distribution of each phylum was found to be driven by 4 to 10 parameters. The drivers can be ranked in the following sequence, according to their cumulated influence on all phyla: pH > land management > soil texture > soil nutrients > climate. Soil pH was the major driver and significantly explained the variation of 17 phyla with a maximum of 27.3% for Bacteroidetes. The influence of an increasing pH was positive for nine phyla and negative for eight phyla (Fig. 6B). Regarding land management influence, eight phyla were stimulated and six were inhibited by agricultural practices along a gradient from forest to vineyards (Fig. 6C). Soil texture (clay and silt contents) was a driver for 17 phyla, mainly minors, and explained up to 6% of their variance (Fig. 6B). Considering the soil nutrient characteristics [C/N ratio, soil organic carbon (SOC), available phosphorus, and total potassium], most phyla were significantly but weakly influenced by at least one parameter, with less than 4% of the explained variance (Fig. 6B). More precisely, the C/N ratio significantly influenced nine phyla, followed by SOC, available phosphorus, and total potassium. For climate, temperature was the only significant but weak driver for seven phyla, inhibiting six of these but stimulating Crenarchaeota (Fig. 6B). Finally, spatial descriptors and especially PCNM vectors, summarized here in the fine (30 to 100 km) and coarse (100 to 350 km) spatial scale effects, affected all phyla except Alphaproteobacteria (Fig. 6D).

At the genus level, the total explained variance in taxa distribution ranged from 13.2 to 72.2% and appeared greater than that of the phyla (table S5). According to the amounts of variance explained by the different environmental parameters, the drivers can be ranked in the following sequence: pH > soil nutrients > land management > spatial descriptors. Soil texture and climate were only minor drivers even though their effects were statistically significant for half of the genera. As for the phylum level, soil pH was the major driver for 30 genera, with an average of 14.0% of the explained variance and a maximum of 50.9% for *Isosphaera* (Planctomycetes). Similarly, soil

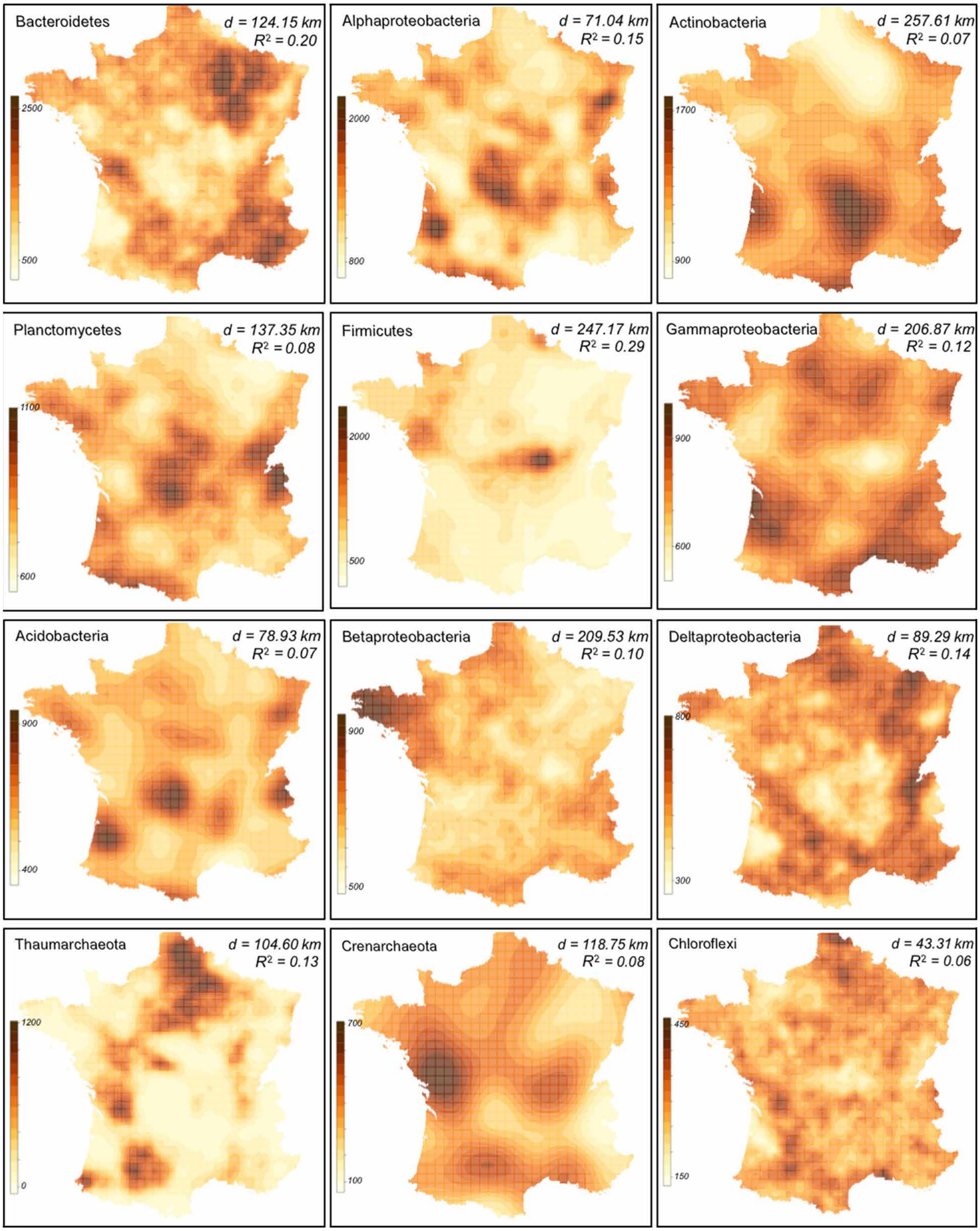

Fig. 4. Mapping of abundance of the most dominant bacterial and archaeal phyla across France. For each map, *d* is the range in kilometers estimated by the model, and R^2 corresponds to the correlation between the predicted and measured values. The quality parameters and the model types are detailed in table S2.

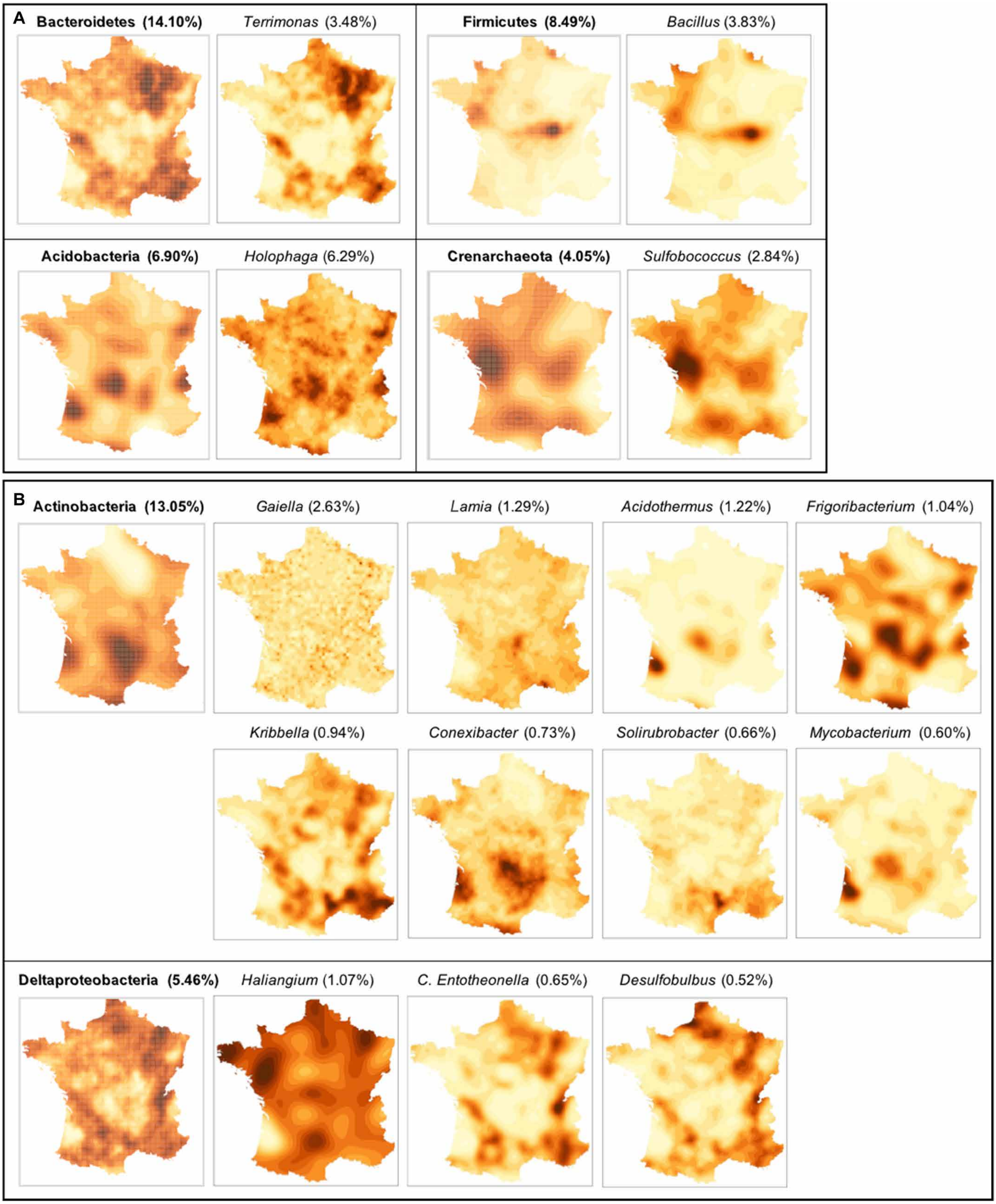

Fig. 5. Comparison of maps phylum/genera. (**A**) Four examples of a phylum for which the spatial distribution is consistent with its major genus. (**B**) Two examples of a phylum for which the spatial distribution represents the cumulative distributions of all genera belonging to the phylum. The complete set of maps is available in fig. S4.

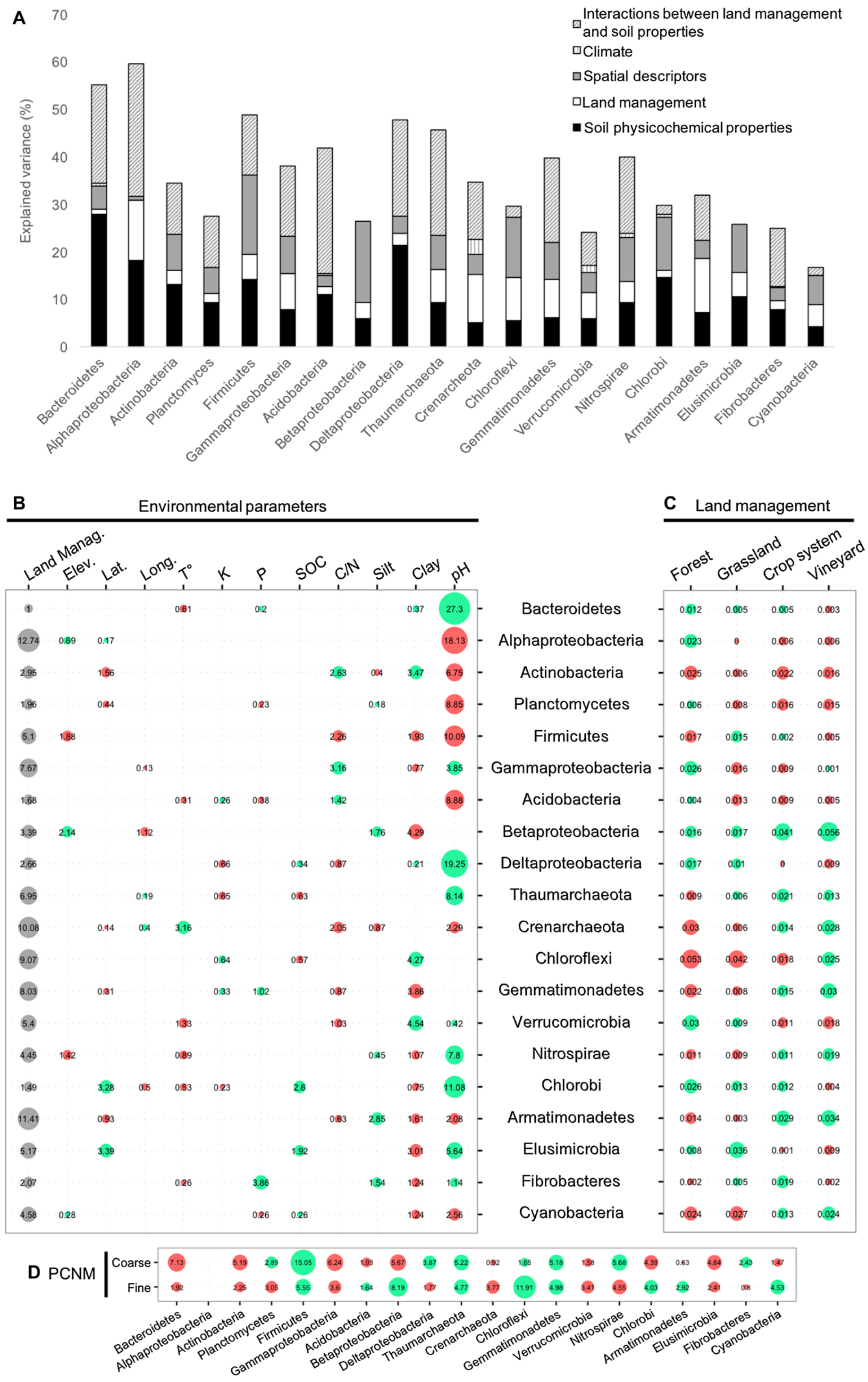

Fig. 6. Variance partitioning of the microbial phyla across France according to environmental and spatial parameters. (**A**) The 20 microbial phyla are ranked from the most to the least abundant. The explained variance corresponds to the sum of the adjusted R^2 values of the significant parameters within the contextual groups (soil physicochemical parameters, land management, spatial descriptors, climate, interactions between soil physicochemical properties and land management). The threshold for statistical significance was set at 0.01. Missing values indicate that no variable of the related group was retained in the model. (**B** to **D**) Contribution and effect of environmental parameters, land management, and spatial descriptors (PCNM at medium and coarse scales) on the distribution of bacterial and archaeal phyla. The colors depict the direction of the standardized partial regression coefficients (green, positive effect; red, negative effect). The height of the shape and the values indicate the percentage of variance explained by environmental parameters [for (B) and (D), proportions are comparable between boxes] and the coefficient of the standardized partial regression of each land management type. For this last effect, the coefficients are relative to a reference level grouping of 60 samples unclassified in the four types (C). The explained variance represents the respective significant contribution of each variable and was calculated by considering all other variables using partial regression models and adjusting the R^2 values.

nutrients (C/N ratio, SOC, available phosphorus, and potassium) were an important driver for soil genera [average = 2.3%, maximum = 9.6% for *Azospira* (Betaproteobacteria)]. Land management and soil texture were the weaker drivers at the genus level. In certain cases, drivers were identified at the genus level but not at the phylum level or inversely (table S5).

DISCUSSION

Here, a comprehensive approach was applied to produce one of the most intensive distribution maps of bacterial and archaeal taxa (phyla and genera) on a broad scale and to identify the ecological processes and environmental drivers regulating their biogeographic variations.

Spatial distribution of bacterial and archaeal phyla

Regarding the ubiquity and dominance of the phyla detected in French soils, the most ubiquitous phyla (Bacteroidetes, Alphaproteobacteria, Actinobacteria, Planctomycetes, Firmicutes, Gammaproteobacteria, Acidobacteria, and Betaproteobacteria) were generally the most abundant, as previously stated (*17*). Two hypotheses already formulated can explain this statement: (i) These abundant microorganisms are easier to detect with our technical procedure, or (ii) potential dispersal is greater for microorganisms with a large population size. With a relative abundance exceeding 5%, these phyla were already reported to be dominant phyla in other soil studies (*7*) and also in marine sediments (*18*), oceans (*19*), and mammalian gut microbiota (*20*). These results reinforce the hypothesis of microbial cosmopolitanism initially based on the postulate of Baas Becking (1934), "Everything is everywhere, but the environment selects" (*21*), (*21*) and reexamined more recently (*22*). Such cosmopolitanism may be partly explained by the different abilities of microorganisms: Alphaproteobacteria, Actinobacteria, Bacteroidetes, and Betaproteobacteria are dispersed by aerosolized soil dusts (*23*) and successfully colonize new environments. Firmicutes and Actinobacteria form resistant physiological stages that allow them to survive in hostile environments (*24*, *25*). Some phyla [for example, Firmicutes (*26*)] are generalists for habitat and substrate. In contrast, minor and rare phyla seem to be less cosmopolitan, which could be related to more restricted ecological niches (*11*), and/or limited abilities to migrate (for example, efficient barriers for dispersal), and/or high rates of active losses due to predation or viral lysis (*27*).

Mapping the 20 most representative phyla systematically revealed a heterogeneous and specific distribution for each phylum. The spottiest distributions, seen in Chloroflexi, Fibrobacteres, and Cyanobacteria, may result from the influence of local filters such as landscape configuration or land management variations (fig. S4) (*4*, *28*), whereas distributions in larger patches (for example, Actinobacteria, Firmicutes, Gammaproteobacteria) may be explained by the presence of large natural barriers (mountain, sea, etc.; http://eusoils.jrc.ec.europa.eu/projects/lucas), the main soil types, and climatic conditions (*28*). For example, the hot spots of Acidobacteria located in the southwest closely corresponded to the most acidic soils in France (fig. S4), as also observed in American soils (*8*). In addition, the hot spots of Actinobacteria recorded in Landes (southwest France) and Centre (central France) could be related to distribution of particular types of land management, notably forest and grasslands, in these regions (*29*). These contrasting distribution patterns suggest not only that all the studied phyla can be differently affected by the selection process due to the influence of particular environmental parameters but also that minor and rare phyla can be differently influenced by neutral processes, especially dispersal limitations (*30*). The latter depends on the abilities of phyla members to disperse across a large territory through passive or active mechanisms and to persist at the settling location (*23*). Together, the gradient of patch size might be partly due to a gradient of selection by environmental parameters and/or dispersal limitation.

Which processes and drivers for which phyla?

The total explained variance in phyla distribution ranged from 60% (Alphaproteobacteria) to 17% (Cyanobacteria). The latter phylum is known to include photosynthetic organisms mainly found in aquatic environments like oceans or freshwater (*31*) and generally influenced by temperature, light, specific nutrients, and competitor abundances (*32*). Thus, soil moisture and light accessibility, not measured in this study, might be better predictors of the variability of this phylum in soils and justify the weak explained variance observed in our study. The low explained variance of Cyanobacteria may also be due to their restricted ecological niches, which is limited to the soil surface (that is, light exposed) that could be diluted when soils were collected at 0- to 30-cm depth, and in biofilm associated with coarse particles that could be excluded when soils were sieved through a 2-mm mesh.

Since 16 of our phyla were mainly influenced by environmental parameters, our results support a previous assumption (*30*), based on a review of 22 studies, that environmental selection has a greater effect than distance in shaping microbial phyla distribution (*29*). Ranking of the environmental data sets indicated that phyla distribution was more dependent on local selection processes like soil properties and land management. The impact of climate could be masked by interactions between soil, land management, and climatic characteristics, which represented between 0 and 28% of the explained variance and an important effect for 10 phyla (Fig. 6A). This taxa-based hierarchy reinforced the previous results obtained for microbial biomass (*16*), diversity (*3*, *5*), and genetic structure (*33*) of the microbial community.

Spatial descriptors mainly explained the variation of four phyla. Thus, as previously shown by our mapping, dispersal ability depended on the phylum (*34*) and the ability of that phylum to survive in a new location. As mentioned above, the potential of Betaproteobacteria and Firmicutes to disperse and to colonize soil is high (*23*, *25*). These biological characteristics are consistent with large patches on the maps (>200 km; Fig. 4) and the observed cosmopolitan and abundant distributions of both phyla (Fig. 2). Chloroflexi and Cyanobacteria were also cosmopolitan but less abundant. This finding suggests that these phyla (i) exhibit a weak ability to colonize a wide range of soil types rather than a weak ability to migrate, (ii) may not be able to produce large populations in most soils, or that (iii) their detection signal is diluted due to inappropriate soil sampling and sieving strategy (as described below).

Among the environmental parameters, soil pH is the major driver for 17 phyla, which confirmed the overriding effect of pH on the microbial community as a whole (*1*). However, our results contradict some other reports on pH effects on the dominant phyla. For example, the negative influence observed for Actinobacteria is in accordance with some studies (*8*) but not with others (*35*). These discrepancies in pH effect might be due to different interactions between the soil properties producing soil heterogeneity. In addition, dissimilar results could also be due to the low sampling effort and a priori sampling strategy applied in most studies, which limited soil variations and therefore the genericity of the results (*33*, *36*), and to the technical approaches involved (cultivable versus molecular), which could bias the estimation of taxa abundance and pH effect (*8*).

Across France, the significant influence of land management was linked to cropping intensity and associated soil perturbation with a gradient from forests to vineyards/orchards (*16*). For example, in our study, Alphaproteobacteria were more abundant in forests, which are known to undergo weak soil perturbation and to provide copiotrophic habitats rich in recalcitrant organic matter (*37*). This observation is in accordance with the ability of several genera belonging to this phylum, like *Rhizobium*, to degrade recalcitrant organic matter, such as plant lignin (*35*). Alphaproteobacteria are also known to have important nitrogen cycle functions and to be involved in the decomposition of lignin by-products (mono- and oligophenolic compounds) relevant to forest ecosystems. The inorganic fertilization of soils under forest-to-agroecosystem conversion may also explain the low abundance of nitrogen-fixing Alphaproteobacteria (*38*).

The effect of fine soil texture was negative for 10 phyla, with silt and clay having a differential impact in certain cases (Actinobacteria, Betaproteobacteria, Nitrospirae, Armatimonadetes, and Fibrobacteres). This pattern suggested that half of the bacterial and archaeal phyla were adapted to coarse-textured soils, considered as heterogeneous, less-protected, and oligotrophic habitats (*39*). In addition, the weakness of the observed effect of soil nutrients (C/N ratio, SOC content, available phosphorus, and total potassium) suggests that microbial phyla cannot be classified into copiotrophic or oligotrophic categories based on soil nutrient characteristics alone (*8*). Finally, the weak but significant positive effect of climate and, especially, temperature on Crenarchaeota (3.2% of explained variance) is consistent with the ability of several genera belonging to this phylum to survive in high-temperature habitats such as hydrothermal vents or volcanoes (*12*).

The significant part of the variance explained for four phyla by spatial descriptors and, especially, PCNM vectors might be partly related to (i) variations in unmeasured environmental parameters at different spatial scales (*30*) and/or (ii) the dispersal ability, as a neutral process, of the phyla (*40*). On the basis of the latter postulate, three response patterns were observed, depending on the scale of spatial structure, which led to three hypotheses for the dispersal abilities of microorganisms: (i) positive effects of both fine and coarse scales on four phyla, suggesting their dispersal over short and long distances; (ii) positive effects of only fine scale on five phyla, suggesting their dispersal over short distance; and (iii) negative effects of both scales on six phyla, suggesting high dispersal limitation. The coarse-scale effect was the strongest (15%) and positive for Firmicutes, while the fine scale had the highest and positive effect on Chloroflexi (11.9% of variance). These observations accord with our previous findings that these variations were mainly explained by spatial descriptions and suggested that the distribution of these phyla is greatly influenced by dispersal. They also accord with our mapping results since Firmicutes were structured in large spatial patches (>200 km), whereas Chloroflexi were spatially distributed in small patches (<60 km). All our hypotheses about PCNM vectors and dispersal need to be validated by computing a complementary approach using a neutral model of metacommunity (*41*).

Disentangling phylum-level biogeography from genus-level ecology

To investigate the ecological coherence of phyla biogeography, we also examined the spatial distribution and environmental drivers of the main genera detected in French soils and compared the results between genus and phylum levels. Depending on the phylum, spatial distribution and environmental drivers were consistent with the dominant genus in the phylum or with the cumulative distributions and drivers of the genera within the phylum.

The first scenario was exemplified by the phyla Acidobacteria, Firmicutes, and Bacteroidetes, whose overall distribution perfectly matched that of their dominant genus, that is, *Holophaga* (on average 6.3% of total sequences), *Bacillus* (3.8%), and *Terrimonas* (3.5%), respectively, for both spatial distribution and environmental drivers. The distribution of *Holophaga* overlaid the soil pH map (fig. S4), confirming the high affinity of Acidobacteria for low pH (*8*). The environmental drivers of *Holophaga* indicated that the Acidobacteria members inhabited constrained soil environments, that is, acidic and oligotrophic soils with a coarse texture. According to the sparse knowledge of this uncultivable genus (one species known), these bacteria have the ecological role of K-strategists, with a strictly anaerobic metabolism, capacities for homoacetogenesis, and also the genomic potential for involvement in key N-cycle processes such as nitrogen fixation (*42*). Thus, the high abundance of *Holophaga* suggests the presence of a nonnegligible amount of anoxic microhabitats (micropores) in soils, characterized by low nutrient availability within the soil aggregates (*43*).

Similar drivers were identified for the phylum Firmicutes and genus *Bacillus*, such as soil pH, clay content, C/N ratio, and elevation. The negative effect of all these environmental parameters indicated that *Bacillus* inhabit cropland and grassland soils with high pH and coarse texture rich in labile organic matter. This is in accordance with the ability of *Bacillus* to degrade simple organic compounds such as xylose (*44*) or starch (*26*), commonly found in agricultural soils. Moreover, most *Bacillus* species (347 referenced to date) are phosphate solubilizers in the soil (*45*) and carry the genes required for six nitrogen cycle pathways [from ammonia assimilation to nitrogen fixation (*46*)]. This strong involvement of *Bacillus* in the biogeochemical cycles, which subsequently increases the availability of nutrients to plants, together with its capacity for phytopathogen suppression (*47*), confers this genus with a key position in soils, especially in the cropping context.

The Actinobacteria and Gammaproteobacteria phyla provided an example of the second scenario. The ecology and distribution of each of these phyla reflects the combined ecological and biogeographic characteristics of all the genera in that phylum. Members of Actinobacteria are known to have an important role in organic matter turnover and the breakdown of recalcitrant molecules, such as cellulose and complex hydrocarbons, and are particularly abundant in woodland soils where the C/N ratio is highest (*48*). In our data set, the genera *Frigoribacterium*, *Acidothermus*, *Conexibacter*, and *Mycobacterium* were more abundant in forests or correlated with a high C/N ratio. Moreover, as numerous Actinobacteria genera are sources of antibiotics, insecticides, and antifungal or bioherbicide agents, they have a key role in biological methods of crop protection (*49*). These genera also include plant growth–promoting rhizobacteria, symbionts, endophytes, and elicitors of plant defense (*49*); thus, several of them should find appropriate ecological niches in crop systems. All these soil functions, added to potential implications in nitrogen cycling (*46*), may explain why distribution of the Actinobacteria phylum is so widespread.

Robustness and limitation of the sampling and analytical strategies

This work is the most intensive, without a priori, soil sampling survey (about 2000 soil samples) focusing on a nationwide scale. Compared to global studies (on the world scale), based on a few tens of samples (*14*), our sampling design provides a foundation for robust analysis and conclusions about soil microbial biogeography. By examining major environmental variability across a 550,000 km^2 area (*5*), we can affirm that our results are not biased by environmental sampling (*50*). The general applicability of our results to other pedoclimatic and land use conditions (for example, tropical ecosystems) may be limited by the small surface area of France in relation to that of the world. However, the pedoclimatic diversity recorded on this regional scale is higher than in many other countries (*51*). This diversity in turn provides a wide range of environmental conditions for microbial communities that can also be found in a large part of the continental Europe and, more globally, in the Northern Hemisphere (*51*).

Our molecular analytical strategy is known to present high levels of robustness (*52*), although numerous biases inherent in amplicon library preparation like DNA extraction (*53*), amplification (*54*), sequencing, and the inference of patterns of organism abundance from library data pertaining to relative abundance are also well known (*55*). Analyses were conducted in a consistent manner to remove errors due to sequencing and chimeras, and the data sets were rarefied to the same sampling depth (that is, 10,000 reads per sample), so that relative changes in microbial taxonomic composition levels can be compared across samples, even if the biodiversity sampling was not exhaustive (*55*). Nevertheless, comparison of our study with others is limited by our choice of 16*S* rRNA primers, which were designed to specifically target both bacterial and archaeal diversity. The results obtained by using these primers were able to reveal taxonomic groups (for example, *Holophaga*), rarely detected in previous soil studies. Several of these taxa, mainly genera, occurred frequently and in high relative abundance on the nationwide scale, implying that their detection was not random. In addition to using these different primers, we also chose a finer but more time-consuming method of taxonomic assignment than the approach currently used in QIIME (Quantitative Insights Into Microbial Ecology). More precisely, rather than assigning the seed sequence of each operational taxonomic unit (OTU), all reads in the data set were individually assigned. This approach led to changes in relative abundance of taxonomic groups within the community and, thus, the detection of new phyla or genera, underestimated in other studies.

CONCLUSION

Our study highlights the heterogeneous distribution of all soil bacterial and archaeal phyla and genera across continental France, using one of the most intensive soil sampling strategies available. Biogeographical patterns ranged from patchy to spotty and were explained by both selection and neutral processes, each being nonexclusive for a given phylum. Comparison of our wide-scale study with investigations conducted on worldwide, regional, and landscape scales indicated that soil pH and land management are recurrent drivers (*10*, *14*, *36*). Comparison of bacterial biogeography at the phylum and genus taxonomic levels suggests that analysis at high taxonomic levels is mainly suitable for deciphering the distribution and environmental drivers of dominant populations. Comprehensive knowledge of the ecological attributes and spatial distribution of soil bacteria should improve the ability to predict shifts in community structure and, therefore, in soil functioning. Finally, on the basis of this knowledge, the future objectives will be to increase soil management sustainability and to implement the corresponding protection policy in a context of global change. Here, a significant amount of unexplained variance was observed for most of the soil bacterial taxa distributions. A promising way to better decipher the drivers behind this residual variance would be to analyze bacterial taxa by soil horizon to detect the potential effects of horizon-specific soil properties such as pH or C/N ratio. It is also important to note that the hierarchy of the processes and drivers was mainly based on abiotic parameters. Hence, another perspective could be to explore bacterial and archaeal biogeography in light of the biotic relationships existing between soil populations through interaction networks (*2*). Although interactions between community members with regard to functioning are undoubtedly huge, the identification and integration of these biotic and abiotic interactions on a broad scale still present major challenges in microbial ecology. In addition, consideration of the fungal populations will further enhance our global overview of soil microbial taxonomic group distribution by integrating the biotic interrelationships existing between these two kingdoms of soil microorganisms.

MATERIALS AND METHODS

Experimental design

Soil samples were obtained from the French Soil Quality Monitoring Network (RMQS), which is a soil monitoring network based on a 16-km regular grid across the 550,000-km^2 French territory (*5*, *16*). The RMQS includes 2173 monitoring sites, each located at the center of a 16 × 16–km cell (Fig. 1), for which soil profile, site environment, climatic factors, location, vegetation, and land management were described. All details concerning soil sampling, storage, and physicochemical analysis were reported (*16*). Available climatic data were monthly rain, ETP (evapotranspiration), and temperature at each node of a 12 × 12–km^2 grid, averaged for the 1992–2004 period. These climatic data were obtained by interpolating observational data using the Analysis System Providing Information Adapted to Nivology (SAFRAN model). The RMQS site–specific data were linked to the climatic data by finding the closest node within the 12 × 12–km^2 climatic grid for each RMQS site. Land cover was recorded according to the coarse level of the CORINE Land Cover classification (IFEN; www.statistiques.developpement-durable.gouv.fr/donnees-ligne/li/2539/0/base-donnees-geographique-corine-land-cover-clc.html), which consists of a rough descriptive classification into five classes: forest, croplands, grasslands, others, and perennial crops (corresponding to vineyards and orchards). All these data were available in the DONESOL database (*16*, *28*).

Molecular characterization of bacterial community diversity

Soil DNA extraction and purification

Microbial DNA was extracted and purified from 1 g of each of the 2173 soils sampled at each RMQS site, using the GnS-GII procedure previously described (*53*). Crude DNA extracts were quantified by agarose gel electrophoresis stained with ethidium bromide and using calf thymus DNA as standard curve (*16*). Crude DNA was then purified using a MinElute gel extraction kit (Qiagen) and quantified using a QuantiFluor staining kit (Promega) prior to further investigations.

Polymerase chain reaction amplification and pyrosequencing of 16S rRNA gene sequences

A 16*S* rRNA gene fragment targeting the V3 to V4 regions to characterize bacterial diversity was amplified using the primers F479 (5′-CAGCMGCYGCNGTAANAC-3′) and R888 (5′-CCGYCAATTCMTTTRAGT-3′) with the method described previously (*53*). Homemade bioinformatic programs were developed to design these primers and to search large DNA sequence databases for the presence of primers, including degeneracies, as coded by the IUPAC (International Union of Pure and Applied Chemistry) rules, and also additional mismatches to test primer improvement. The sequences investigated were SILVA, direct extraction of every small subunit rRNA sequence from EMBL (European Molecular Biology Laboratory) using ACNUC, and also a dedicated reference database of 18*S* eukaryotic sequences, which had been thoroughly analyzed and annotated for in silico match analysis [see table S2 from Terrat *et al.* (*7*)]. A total of 2132 soil samples were successfully amplified from the 2173 DNA soil samples. The polymerase chain reaction (PCR) products were then purified using a MinElute PCR purification kit (Qiagen) and quantified using a QuantiFluor staining kit (Promega). A second PCR of seven cycles was then duplicated for each sample under similar PCR conditions, with purified PCR products as matrix (7.5 ng of DNA was used for a 25-μl mix of PCR) and dedicated fusion primers ("F479/AdaptorB," "R888/MID/AdaptorA") integrating the required adaptors, keys, and multiplex identifiers at the 5′ extremities. All duplicated PCR products were then pooled, purified using a MinElute PCR purification kit (Qiagen), and quantified using a QuantiFluor staining kit (Promega). For all libraries, equal amounts from 30 samples were pooled and then cleaned to remove excess nucleotides, salts, and enzymes using the Agencourt AMPure XP system (Beckman Coulter Genomics). TE buffer (100 μl) (Roche) was used for the elution. Pyrosequencing was then carried out on a GS FLX Titanium (Roche 454 Sequencing System) by Genoscope.

Bioinformatics sequence analysis

Bioinformatic analyses were done using the GnS-PIPE pipeline developed by the GenoSol platform [Institut National de Recherche Agronomique (INRA)] (availability: https://zenodo.org/record/1123425#.WxD4DO6FPIU) (*5*). After sequencing, 49,794,516 raw reads were obtained for the 2132 soil samples. First, all raw reads were sorted according to each multiplex identifier sequence (no tolerated error in 10-base multiplex identifiers). Then, a preprocessing step was carried out to filter and delete low-quality reads based on (i) their length (fewer than 350 bases), (ii) their number of ambiguities (deletion of reads with one or more N, or reads with homopolymer of more than seven consecutive bases), and (iii) their primer sequence(s) (the proximal primer sequence had to be complete and without errors, with a maximum of two mismatches tolerated in the distal primer sequence). A PERL program was then applied for rigorous dereplication (that is, clustering of strictly identical sequences with same length). The dereplicated reads were aligned using the INFERNAL alignment program (v1.0.2, http://eddylab.org/infernal, with the selected parameters: --hbanded, --sub, --dna) to obtain a global alignment against a hand-curated database containing 508 full-length 16*S* rRNA sequences, chosen after careful consideration and aligned with recommended parameters (v1.0.2 using selected parameters: --rf, --ere 1.4). Then, in agreement with a previous study (*56*), aligned sequences were clustered into OTUs at 95% of similarity using the CrunchClust program (v43 program, https://code.google.com/archive/p/crunchclust) that groups rare reads with abundant ones and does not count differences in homopolymer lengths (default parameters were selected). We chose this level of clustering as it corresponds roughly to the genus level, particularly with our primer set (in silico evaluation), and it also corresponds to the level used to define each soil community composition. Here, an OTU is defined by the most abundant read, known as the centroid, and every read in the OTU must have similarity above the given identity threshold with the centroid. A filtering step was then carried out to check all reads detected only once and not clustered, which might be artifacts, such as PCR chimeras, based on the quality of their taxonomic assignments. A database of 16*S* rRNA sequences from SILVA (version r114), filtered, curated, and annotated (database is available here: https://zenodo.org/record/1065438#.WxD4L-6FPIU) using the USEARCH program (v8.0.1623; www.drive5.com/usearch) with specific parameters (-maxhits 15, -maxaccepts 0, and –maxrejects 0), was used (associated PERL program using USEARCH and formatting results are available here: https://zenodo.org/record/1064170#.WxD4S-6FPIU). If a read obtained a percentage of similarity lower than 90%, then it was discarded from the data set. This filtering step allowed the deletion of most of the chimeras produced during the PCR process, but also led to the deletion of potential novel minor taxa not currently included in the used database. A total of 32,634,692 high-quality reads (range of sequencing depth: 48 to 49,926 reads by sample) were kept after these steps. The number of high-quality reads for each sample was then "rarefied" (that is, 10,000 high-quality reads for each sample) by random selection to allow efficient comparison of the data sets and avoid biased community comparisons and rarefaction curves (*5*). So, 1798 soil samples were kept for subsequent analyses, encompassing a total of 17,980,000 reads.

A postprocessing step was then applied to this global data set to filter potentially artifactual reads as already described regarding microbial richness across France (*5*). Briefly, the 17,980,000 reads from all samples were aligned and clustered at 95% of similarity into OTUs, using the previously described CrunchClust program. Thereafter, all OTUs that occurred only once in the overall data set and encompassed only a single read were removed. This postprocessing step reduced the number of total OTUs from 205,590 to 92,571 (more than 50% lost), but the number of reads only from 17,980,000 to 17,866,981 (less than 1% lost). The number of deleted reads by sample was 62 ± 60 on average (minimum, 10; maximum, 1093). Finally, all kept reads were then compared to the dedicated reference database originated from SILVA. The same previously described programs and parameters were used to determine independently the composition of each soil community at the phylum or subphylum level (that is, Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria, and Epsilonproteobacteria) and at the genus level (procedure, database, and programs available online: https://zenodo.org/record/1065438#.WxD4fe6FPIU and https://zenodo.org/record/1064170#.WxD4ku6FPIU). Unknown sequences represented an average of 0.47% by sample at the phylum taxonomic level and an average of 11% by sample at the genus taxonomic level.

All raw data sets are publicly available in the European Bioinformatics Institute (EBI) database system (in the Short Read Archive) under project accession no. PRJEB21351. The matrix of taxonomic data is available online (https://zenodo.org/record/1063503#.WxD4pu6FPIU).

Statistical analysis

Phylum and genus classification

On the basis of the average relative abundance and occurrence of the 35 phyla, an ascendant hierarchical clustering was performed using the unweighted pair group method with arithmetic mean (UPGMA) agglomeration method. Despite the seven clusters and the six cutoffs identified by the clustering (fig. S1), we limited the ranking to four groups to keep the classification simple and readable. The subsequent mathematical analyses excluded the 15 rarest phyla, that is, those with a relative abundance below 0.01% and detected in less than 50% of the sampling sites, because the robustness of the statistical analyses was affected by the reduced size of the data set.

The dominant genera were considered present in more than 75% of the sampling sites and with an average relative abundance higher than 0.5% of the sequences. These thresholds were chosen to retain those genera that were most representative at the community level, that is, the pool of dominant genera should represent most of the sequences in each sample.

Mapping using geostatistics

A geostatistical method was used to map the microbial phyla and genera and to characterize their spatial variations (https://zenodo.org/record/1063500#.WxD4t-6FPIU). When the relative abundance of the taxa did not follow a normal distribution, a quantile transformation was applied before modeling the spatial correlations. In conventional geostatistical analysis, an estimate of a variogram model is computed on the basis of the observations, which describe the spatial variation of the property of interest. This model is then used to predict the property at unsampled locations using kriging (*57*). A common method for variogram estimation is first to calculate the empirical (so-called experimental) variogram by the method of moments (*58*), and then to fit a model to the empirical variogram by (weighted) nonlinear least squares. We tried to fit several models and retained the one that minimized the objective function (*59*). The validity of the best fitted geostatistical model was then assessed in terms of the standardized squared prediction errors (SSPEs) using the results of a leave-one-out cross validation. If the fitted model provided a valid representation of the spatial variation of the taxa relative abundance, then these errors would have a χ^2 distribution with a mean of 1 and a median of 0.455 (*60*). The mean and median values of the SSPEs were also calculated for 1000 simulations of the fitted model to determine the 95% confidence limits and to obtain a map of the kriging SE. The geostatistical analysis gstat package was used for variogram analysis and kriging. The effective range of the variograms fitted on the data represents the size of the geographical patches, which were classified as spotty, intermediate, or large spatial patterns. Three classes of patch size were defined according to our sampling procedure across a 16 × 16–km grid. The spottiest patch size, <64-km radius, corresponded to a local spatial pattern considering a spatial structure spread over 51 sampling sites only. The largest patch size, >200-km radius, was considered the more global spatial pattern at the scale of France. Patches of 200-km radius (=400-km diameter) represented more than half of the territory and contrasted with the spottiest patches. The third class grouped all the intermediate patches, due to the lack of information, allowing better ranking of these spatial structures.

Variance partitioning

The relative contributions of soil physicochemical parameters, land management, climatic conditions, geomorphology, and space in shaping the patterns of soil bacterial phyla (called "marginal effects") were estimated by variance partitioning (https://zenodo.org/record/1063479#.WxD4zu6FPIU). The explanatory variables were selected to reduce the autocorrelation in the models and to obtain the most parsimonious models. Thus, only 12 environmental parameters of the 21 measured were kept in the analyses. These were pH, amounts in clay and silt for the soil texture, C/N ratio, SOC, available phosphorous and total potassium content for the soil nutrient, temperature for the climatic factors, latitude and longitude of the location of the sites, elevation for the geomorphology, and four classes along an anthropization gradient (forest, grassland, crop system, and vineyard/orchard) for the land management. All quantitative (response and explanatory) data were standardized to guarantee an approximated Gaussian and homoskedastic residual distribution of the model. Considering that the largest part of the environmental selection was measured by the previous explanatory variables, we then investigated the effect of neighborhood on the residuals of the variance partitioning models, using the PCNM approach computed on the spatial coordinates of the sites. The PCNM method (described below) describes and identifies the scales of the spatial relationship between samples (*40*) (see details in the Supplementary Materials). The classification of the PCNM vectors into fine-scale (30 to 100 km) and coarse-scale (100 to 350 km) neighborhood effects was chosen to reveal those spatial structures in the residual distributions of phyla, which could be related to short- or long-distance dispersal patterns. Thereafter, the most influential types of parameters were identified by organizing the above parameters into five groups: (i) soil physicochemical characteristics including pH, soil texture, and nutrient contents; (ii) land management; (iii) spatial descriptors including spatial location, elevation, and PCNM variables; (iv) climate as temperature; and, finally, (v) interactions between all the environmental parameters excluding the PCNM variables. The variance explained by each group of parameters was computed as the sum of the variance explained by all marginal effects.

The PCNM approach was used to describe and identify the scales of the spatial relationship between samples (*40*). This PCNM method was applied to the geographic coordinates, and only PCNMs with a significant Moran index were selected for the variance partitioning analysis ($P < 0.001$). These PCNMs represented the spatial scales that the sampling scheme could perceive (*61*). The spatial neighborhood described by each PCNM was determined by the range of Gaussian variogram models (*62*). All quantitative (response and explanatory) data were standardized to have an approximated Gaussian and homoskedastic residual distribution. A two-step procedure was used to determine the environmental parameters significantly shaping bacterial phyla and to limit overfitting and exclude co-linear variables (*63*). The first step consisted of a coarse selection of explanatory variables included in models minimizing the Bayesian Information Criterion and maximizing the adjusted R^2 using the regsubset function (leaps package). In the second step, a forward selection procedure was applied to the subset of explanatory variables to identify the model maximizing the adjusted R^2 (*63*). Spatial descriptors were then selected from the model residuals (*64*) using the forward selection step only since all PCNMs are linearly independent. The respective amounts of variance (that is, marginal and shared) for bacterial phyla distribution were determined by canonical variation partitioning and the adjusted R^2 with redundancy analysis (*63*). The statistical significance of the marginal effects was assessed from 1000 permutations of the reduced model. All these analyses were performed with R (www.r-project.org/) using the vegan package (https://cran.r-project.org/web/packages/vegan/vegan.pdf).

The relative contributions of environmental parameters were evaluated by performing a variance partitioning for each phylum (the applied procedure is detailed in the Supplementary Materials, https://zenodo.org/record/1063479#.WxD43u6FPIU).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/7/eaat1808/DC1

Fig. S1. Dendrogram of the ascendant hierarchical clustering by the UPGMA agglomeration method.

Fig. S2. Mapping of abundance of minor bacterial phyla.
Fig. S3. Mapping of abundance of the 47 major bacterial genera belonging to 12 phyla.
Fig. S4. Maps of soil physicochemical characteristics across France.
Table S1. Comparisons of relative abundance distribution between the 35 phyla across France.
Table S2. Parameters and type of models for the 20 maps of phyla.
Table S3. Outputs of models used for the variance partitioning analysis of phyla distributions.
Table S4. Occurrence, relative abundance, and parameters of geostatistical models for the 47 genera.
Table S5. Outputs of models used for the variance partitioning analysis of genera distributions.

REFERENCES AND NOTES

1. R. I. Griffiths, B. C. Thomson, P. Plassart, H. S. Gweon, D. Stone, R. E. Creamer, P. Lemanceau, M. J. Bailey, Mapping and validating predictions of soil bacterial biodiversity using European and national scale datasets. *Appl. Soil Ecol.* **97**, 61–68 (2016).
2. B. Karimi, P. A. Maron, N. Chemidlin Prévost-Bouré, N. Bernard, D. Gilbert, L. Ranjard, Microbial diversity and ecological networks as indicators of environmental quality. *Environ. Chem. Lett.* **15**, 265–281 (2017).
3. R. I. Griffiths, B. C. Thomson, P. James, T. Bell, M. Bailey, A. S. Whiteley, The bacterial biogeography of British soils. *Environ. Microbiol.* **13**, 1642–1654 (2011).
4. L. Ranjard, S. Dequiedt, N. Chemidlin Prévost-Bouré, J. Thioulouse, N. P. A. Saby, M. Lelievre, P. A. Maron, F. E. R. Morin, A. Bispo, C. Jolivet, D. Arrouays, P. Lemanceau, Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat. Commun.* **4**, 1434 (2013).
5. S. Terrat, W. Horrigue, S. Dequiedt, N. P. A. Saby, M. Lelièvre, V. Nowak, J. Tripied, T. Régnier, C. Jolivet, D. Arrouays, P. Wincker, C. Cruaud, B. Karimi, A. Bispo, P. A. Maron, N. Chemidlin Prévost-Bouré, L. Ranjard, Mapping and predictive variations of soil bacterial richness across France. *PLOS ONE* **12**, e0186766 (2017).
6. L. R. Thompson, J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, J. A. Navas-Molina, S. Janssen, E. Kopylova, Y. Vázquez-Baeza, A. González, J. T. Morton, S. Mirarab, Z. Z. Xu, L. Jiang, M. F. Haroon, J. Kanbar, Q. Zhu, S. J. Song, T. Kosciolek, N. A. Bokulich, J. Lefler, C. J. Brislawn, G. Humphrey, S. M. Owens, J. Hampton-Marcell, D. Berg-Lyons, V. McKenzie, N. Fierer, J. A. Fuhrman, A. Clauset, R. L. Stevens, A. Shade, K. S. Pollard, K. D. Goodwin, J. K. Jansson, J. A. Gilbert, R. Knight; Earth Microbiome Project Consortium, A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
7. F. Constancias, N. P. A. Saby, S. Terrat, S. Dequiedt, W. Horrigue, V. Nowak, J.-P. Guillemin, L. Biju-Duval, N. Chemidlin Prévost-Bouré, L. Ranjard, Contrasting spatial patterns and ecological attributes of soil bacterial and archaeal taxa across a landscape. *Microbiologyopen* **4**, 518–531 (2015).
8. N. Fierer, M. A. Bradford, R. B. Jackson, Toward an ecological classification of soil bacteria. *Ecology* **88**, 1354–1364 (2007).
9. B. Zhang, X. Wu, G. Zhang, S. Li, W. Zhang, X. Chen, L. Sun, B. Zhang, G. Liu, T. Chen, The diversity and biogeography of the communities of Actinobacteria in the forelands of glaciers at a continental scale. *Environ. Res. Lett.* **11**, 054012 (2016).
10. S. M. Hermans, H. L. Buckley, B. S. Case, F. Curran-Cournane, M. Taylor, G. Lear, Bacteria as emerging indicators of soil condition. *Appl. Environ. Microbiol.* **83**, e02826-16 (2017).
11. J. M. DeBruyn, L. T. Nixon, M. N. Fawaz, A. M. Johnson, M. Radosevich, Global biogeography and quantitative seasonal dynamics of *Gemmatimonadetes* in soil. *Appl. Environ. Microbiol.* **77**, 6295–6300 (2011).
12. S. T. Bates, D. Berg-Lyons, J. G. Caporaso, W. A. Walters, R. Knight, N. Fierer, Examining the global distribution of dominant archaeal populations in soil. *ISME J.* **5**, 908–917 (2011).
13. L. Philippot, S. G. E. Andersson, T. J. Battin, J. I. Prosser, J. P. Schimel, W. B. Whitman, S. Hallin, The ecological coherence of high bacterial taxonomic ranks. *Nat. Rev. Microbiol.* **8**, 523–529 (2010).
14. M. Delgado-Baquerizo, A. M. Oliverio, T. E. Brewer, A. Benavent-González, D. J. Eldridge, R. D. Bardgett, F. T. Maestre, B. K. Singh, N. Fierer, A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).
15. R. Angel, M. I. M. Soares, E. D. Ungar, O. Gillor, Biogeography of soil archaea and bacteria along a steep precipitation gradient. *ISME J.* **4**, 553–563 (2010).
16. S. Dequiedt, N. P. A. Saby, M. Lelievre, C. Jolivet, J. Thioulouse, B. Toutain, D. Arrouays, A. Bispo, P. Lemanceau, L. Ranjard, Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. *Glob. Ecol. Biogeogr.* **20**, 641–652 (2011).
17. D. R. Nemergut, E. K. Costello, M. Hamady, C. Lozupone, L. Jiang, S. K. Schmidt, N. Fierer, A. R. Townsend, C. C. Cleveland, L. Stanish, R. Knight, Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* **13**, 135–144 (2011).
18. C. Bienhold, L. Zinger, A. Boetius, A. Ramette, Diversity and biogeography of bathyal and abyssal seafloor bacteria. *PLOS ONE* **11**, e0148016 (2016).
19. S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. d'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis; *Tara* Oceans Coordinators, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
20. G. P. Donaldson, S. M. Lee, S. K. Mazmanian, Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2015).
21. L. G. M. Baas Becking, *Geobiologie of Inleiding Tot de Milieukunde* (W.P. Van Stockum & Zoon, 1934).
22. J. B. H. Martiny, B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Øvreås, A.-L. Reysenbach, V. H. Smith, J. T. Staley, Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
23. A. Barberán, J. Henley, N. Fierer, E. O. Casamayor, Structure, inter-annual recurrence, and global-scale connectivity of airborne microbial communities. *Sci. Total Environ.* **487**, 187–195 (2014).
24. A. Barberán, K. S. Ramirez, J. W. Leff, M. A. Bradford, D. H. Wall, N. Fierer, Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol. Lett.* **17**, 794–802 (2014).
25. M. Bueche, T. Wunderlin, L. Roussel-Delif, T. Junier, L. Sauvain, N. Jeanneret, P. Junier, Quantification of endospore-forming *firmicutes* by quantitative PCR with the functional gene *spo0A*. *Appl. Environ. Microbiol.* **79**, 5302–5312 (2013).
26. W. R. Horwath, The role of the soil microbial biomass in cycling nutrients, in *Microbial Biomass* (World Scientific Europe, 2016), pp. 41–66.
27. P. E. Galand, E. O. Casamayor, D. L. Kirchman, C. Lovejoy, Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22427–22432 (2009).
28. J. Meersmans, M. P. Martin, E. Lacarce, S. De Baets, C. Jolivet, L. Boulonne, S. Lehmann, N. P. A. Saby, A. Bispo, D. Arrouays, A high resolution map of French soil organic carbon. *Agron. Sustain. Dev.* **32**, 841–851 (2012).
29. E. da C Jesus, T. L. Marsh, J. M. Tiedje, F. M. de S Moreira, Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J.* **3**, 1004–1011 (2009).
30. C. Hanson, J. A. Fuhrman, M. C. Horner-Devine, J. B. H. Martiny, Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506 (2012).
31. P. M. D'Agostino, J. N. Woodhouse, A. K. Makower, A. C. Y. Yeung, S. E. Ongley, M. L. Micallef, M. C. Moffitt, B. A. Neilan, Advances in genomics, transcriptomics and proteomics of toxin-producing cyanobacteria. *Environ. Microbiol. Rep.* **8**, 3–13 (2016).
32. Z. I. Johnson, E. R. Zinser, A. Coe, N. P. Mcnulty, E. M. S. Woodward, S. W. Chisholm, Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**, 1737–1740 (2006).
33. J. Liu, Y. Sui, Z. Yu, Y. Shi, H. Chu, J. Jin, X. Liu, G. Wang, High throughput sequencing analysis of biogeographical distribution of bacterial communities in the black soils of northeast China. *Soil Biol. Biochem.* **70**, 113–122 (2014).
34. D. R. Nemergut, S. K. Schmidt, T. Fukami, S. P. O'Neill, T. M. Bilinski, L. F. Stanish, J. E. Knelman, J. L. Darcy, R. C. Lynch, P. Wickey, S. Ferrenberg, Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* **77**, 342–356 (2013).
35. H. Nacke, A. Thürmer, A. Wollherr, C. Will, L. Hodac, N. Herold, I. Schöning, M. Schrumpf, R. Daniel, Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLOS ONE* **6**, e17000 (2011).
36. S. L. O'Brien, S. M. Gibbons, S. M. Owens, J. Hampton-Marcell, E. R. Johnston, J. D. Jastrow, J. A. Gilbert, F. Meyer, D. A. Antonopoulos, Spatial scale drives patterns in soil bacterial diversity. *Environ. Microbiol.* **18**, 2039–2051 (2016).
37. N. Pascault, L. Ranjard, A. Kaisermann, D. Bachar, R. Christen, S. Terrat, O. Mathieu, J. Lévêque, C. Mougel, C. Henault, P. Lemanceau, M. Péan, S. Boiry, S. Fontaine, P.-A. Maron, Stimulation of different functional groups of bacteria by various plant residues as a driver of soil priming effect. *Ecosystems* **16**, 810–822 (2013).
38. D. VanInsberghe, K. R. Maas, E. Cardenas, C. R. Strachan, S. J. Hallam, W. W. Mohn, Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils. *ISME J.* **9**, 2435–2441 (2015).
39. F. Constancias, N. Chemidlin Prévost-Bouré, S. Terrat, S. Aussems, V. Nowak, J.-P. Guillemin, A. Bonnotte, L. Biju-Duval, A. Navel, J. M. F. Martins, P.-A. Maron, L. Ranjard, Microscale evidence for a high decrease of soil bacterial density and diversity by cropping. *Agron. Sustain. Dev.* **34**, 831–840 (2014).
40. S. Dray, P. Legendre, P. R. Peres-Neto, Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Modell.* **196**, 483–493 (2006).
41. J. R. Powell, S. Karunaratne, C. D. Campbell, H. Yao, L. Robinson, B. K. Singh, Deterministic processes vary during community assembly for ecologically dissimilar taxa. *Nat. Commun.* **6**, 8444 (2015).

42. A. M. Kielak, C. C. Barreto, G. A. Kowalchuk, J. A. van Veen, E. E. Kuramae, The ecology of *Acidobacteria*: Moving beyond genes and genomes. *Front. Microbiol.* **7**, 744 (2016).
43. R. Tecon, D. Or, Biophysical processes supporting the diversity of microbial life in soil. *FEMS Microbiol. Rev.* **41**, 599–623 (2017).
44. C. Pepe-Ranney, A. N. Campbell, C. Koechli, S. Berthrong, D. H. Buckley, Unearthing the microbial ecology of soil carbon cycling with DNA-SIP. *bioRxiv* **10**, 33 (2015).
45. K. Mohammadi, Phosphorus solubilizing bacteria: Occurrence, mechanisms and their role in crop production. *Resources Environ.* **2**, 80–85 (2012).
46. M. B. Nelson, A. C. Martiny, J. B. H. Martiny, Global biogeography of microbial nitrogen-cycling traits in soil. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8033–8040 (2016).
47. R. Sheng, D. Meng, M. Wu, H. Di, H. Qin, W. Wei, Effect of agricultural land use change on community composition of bacteria and ammonia oxidizers. *J. Soils Sediments* **13**, 1246–1256 (2013).
48. A. B. de Menezes, M. T. Prendergast-Miller, P. Poonpatana, M. Farrell, A. Bissett, L. M. Macdonald, P. Toscas, A. E. Richardson, P. H. Thrall, C/N ratio drives soil actinobacterial cellobiohydrolase gene diversity. *Appl. Environ. Microbiol.* **81**, 3016–3028 (2015).
49. E. A. Barka, P. Vatsa, L. Sanchez, N. Gaveau-Vaillant, C. Jacquard, H.-P. Klenk, C. Clément, Y. Ouhdouch, G. P. van Wezel, Taxonomy, physiology, and natural products of *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* **80**, 1–43 (2016).
50. Y.-L. Chen, T.-L. Xu, S. D. Veresoglou, H.-W. Hu, Z.-P. Hao, Y.-J. Hu, L. Liu, Y. Deng, M. C. Rillig, B.-D. Chen, Plant diversity represents the prevalent determinant of soil fungal community structure across temperate grasslands in northern China. *Soil Biol. Biochem.* **110**, 12–21 (2017).
51. B. Minasny, A. B. McBratney, A. E. Hartemink, Global pedodiversity, taxonomic distance, and the World Reference Base. *Geoderma* **155**, 132–139 (2010).
52. C. A. Lozupone, R. Knight, Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11436–11440 (2007).
53. S. Terrat, P. Plassart, E. Bourgeois, S. Ferreira, S. Dequiedt, N. Adele-Dit-De-Renseville, P. Lemanceau, A. Bispo, A. Chabbi, P.-A. Maron, L. Ranjard, Meta-barcoded evaluation of the ISO standard 11063 DNA extraction procedure to characterize soil bacterial and fungal community diversity and composition. *Microb. Biotechnol.* **8**, 131–142 (2015).
54. R. D'Amore, U. Z. Ijaz, M. Schirmer, J. G. Kenny, R. Gregory, A. C. Darby, M. Shakya, M. Podar, C. Quince, N. Hall, A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics* **17**, 55 (2016).
55. J. Zhou, Z. He, Y. Yang, Y. Deng, S. G. Tringe, L. Alvarez-Cohen, High-throughput metagenomic technologies for complex microbial community analysis: Open and closed formats. *MBio* **6**, e02288-14 (2015).
56. T. Větrovský, P. Baldrian, The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLOS ONE* **8**, e57923 (2013).
57. R. Webster, M. A. Oliver, *Geostatistics for Environmental Scientists* (John Wiley & Sons, ed. 2, 2007).
58. G. Matheron, Les variables régionalisées et leur estimation : Une application de la théorie des fonctions aléatoires aux sciences de la nature, thesis, Paris, Masson (1965).
59. B. Minasny, A. B. McBratney, The Matérn function as a general model for soil variograms. *Geoderma* **128**, 192–207 (2005).
60. R. M. Lark, Modelling complex soil properties as contaminated regionalized variables. *Geoderma* **106**, 173–190 (2002).
61. A. Ramette, J. M. Tiedje, Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2761–2766 (2007).
62. E. Bellier, P. Monestiez, J.-P. Durbec, J.-N. Candau, Identifying spatial relationships at multiple scales: Principal coordinates of neighbour matrices (PCNM) and geostatistical approaches. *Ecography* **30**, 385–399 (2007).
63. A. Ramette, Multivariate analyses in microbial ecology. *FEMS Microbiol. Ecol.* **62**, 142–160 (2007).
64. D. Borcard, P. Legendre, C. Avois-Jacquet, H. Tuomisto, Dissecting the spatial structure of ecological data at multiple scales. *Ecology* **85**, 1826–1832 (2004).

Acknowledgments: We thank all the soil surveyors and technical assistants involved in sampling the sites. Thanks are also extended to J. Kreplak for the use of computation capacity and D. Warwick for her comments on the manuscript. **Funding:** This study was granted by ADEME (French Environment and Energy Management Agency) and by "France Génomique" through involvement of the technical facilities of Genoscope (project number ANR-10-INBS-09-08). Because of the involvement of the technical facilities at the GenoSol platform of the infrastructure Analyses et Expérimentations sur les Écosystèmes (ANAEE) France, it also received a grant from the French state through the National Agency for Research under the program "Investments for the Future" (reference ANR-11-INBS-0001). RMQS soil sampling and physicochemical analyses were supported by a French Scientific Group of Interest on soils: the "GIS Sol," involving the French Ministry of Ecology, Sustainable Development and Energy (MEEM); the French Ministry of Agriculture (MAP); the French Institute for Forest and Geographical Information (IGN); the Environment and Energy Management Agency (ADEME); the French Institute for Research and Development (IRD); and the National Institute for Agronomic Research (INRA). **Author contributions:** All authors conceptualized the research project. C.J. coordinated the sampling at the territory scale. M.L. and V.N. executed the molecular analyses. P.W. and C.C. contributed to the DNA sequencing. S.T. and S.D. performed the bioinformatic analyses. C.J., D.A., and N.P.A.S. provided the environmental data set. B.K., N.P.A.S., and N.C.-P.B. contributed to the statistical analyses. B.K., S.T., P.-A.M., N.C.-P.B., and L.R. wrote the original draft. B.K., S.T., S.D., N.P.A.S., W.H., C.J., D.A., P.W., C.C., A.B., P.-A.M., N.C.-P.B., and L.R. reviewed and edited the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials or they are available on different database systems: The sequencing raw data sets are publicly available in the EBI database system (in the Short Read Archive) under project accession no. PRJEB21351. The environmental data that support the findings of this study are available from US 1106 Infosol (INRA Orléans, France), but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Additional data related to this paper may be requested from the authors and with permission of US 1106 Infosol (INRA Orléans, France). All other data and procedures are available online: taxonomic data set: https://zenodo.org/record/1063503#.WxD5Le6FPIU; SILVA database available here: https://zenodo.org/record/1065438#.WxD5ae6FPIU); bioinformatic procedure: https://zenodo.org/record/1064170#.WxD5fe6FPIU; procedure for variance partitioning: https://zenodo.org/record/1063479#.WxD5ku6FPIU; procedure for mapping: https://zenodo.org/record/1063500#.WxD5qe6FPIU.

Submitted 2 February 2018
Accepted 23 May 2018
Published 4 July 2018
10.1126/sciadv.aat1808

Citation: B. Karimi, S. Terrat, S. Dequiedt, N. P. A. Saby, W. Horrigue, M. Lelièvre, V. Nowak, C. Jolivet, D. Arrouays, P. Wincker, C. Cruaud, A. Bispo, P.-A. Maron, N. C. Prévost-Bouré, L. Ranjard, Biogeography of soil bacteria and archaea across France. *Sci. Adv.* **4**, eaat1808 (2018).