

ARTICLE

Received 10 Oct 2012 | Accepted 2 Jan 2013 | Published 5 Feb 2013

DOI: 10.1038/ncomms2431

# Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity

L. Ranjard<sup>1,2</sup>, S. Dequiedt<sup>2,\*</sup>, N. Chemidlin Prévost-Bouré<sup>1,\*</sup>, J. Thioulouse<sup>3</sup>, N.P.A. Saby<sup>4</sup>, M. Lelievre<sup>2</sup>, P.A. Maron<sup>1,2</sup>, F.E.R. Morin<sup>2</sup>, A. Bispo<sup>5</sup>, C. Jolivet<sup>4</sup>, D. Arrouays<sup>4</sup> & P. Lemanceau<sup>1</sup>

Spatial scaling and determinism of the wide-scale distribution of macroorganism diversity has been largely demonstrated over a century. For microorganisms, and especially for soil bacteria, this fundamental question requires more thorough investigation, as little information has been reported to date. Here by applying the taxa-area relationship to the largest spatially explicit soil sampling available in France (2,085 soils, area covered  $\sim 5.3 \times 10^5 \text{ km}^2$ ) and developing an innovative evaluation of the habitat-area relationship, we show that the turnover rate of bacterial diversity in soils on a wide scale is highly significant and strongly correlated with the turnover rate of soil habitat. As the diversity of micro- and macro-organisms appears to be driven by similar processes (dispersal and selection), maintaining diverse and spatially structured habitats is essential for soil biological patrimony and the resulting ecosystem services.

<sup>1</sup>INRA, AgroSup Dijon, Université Bourgogne, UMR 1347 Agroécologie, 17, rue Sully, BP 86510, F-21000 Dijon, France. <sup>2</sup>INRA, Platform GenoSol, UMR 1347 Agroécologie, 17, rue Sully, BP 86510, F-21000 Dijon, France. <sup>3</sup>Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France. <sup>4</sup>INRA, US 1106 InfoSol, 2163 Avenue de la Pomme de Pin, F-45075 Orléans, France. <sup>5</sup>ADEME, Service Agriculture et Forêt, 20, Avenue du Grésillé—BP 90406, 49004 Angers Cedex 01, France \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to L.R. (email: ranjard@dijon.inra.fr).

**S**ince the first statement of microbial biogeography developed by Baas Becking<sup>1</sup> in 1934 “Everything is everywhere, but, the environment selects”, it is now well established that soil microbial communities exhibit biogeographical patterns, that is, variations in their distribution across space and time<sup>2</sup>. Although the increasing number of studies of spatial microbial ecology from local to wide scale<sup>3</sup> is helping to identify ecological processes (selection, dispersal, ecological and evolutionary drifts, speciation) impacting microbial diversity, the hierarchy of these processes remains unclear in comparison with that of macroorganisms.

One way to tackle the spatial processing of microbial diversification is to evaluate the empirical relationship between geographical distance and genetic similarity of microbial assembly<sup>4,5</sup>, also known as the distance–decay relationship or the taxa-area relationship (TAR)<sup>6</sup>. Application of the TAR has demonstrated that the species richness of eukaryotic macroorganisms increases with the sampling area<sup>7</sup>. This significant rate of diversity turnover has generally been explained by a dynamic equilibrium between extinction and immigration (“area” hypothesis, mediated by dispersal) and/or by the increase of habitat diversity with increasing area (“habitat heterogeneity” hypothesis, mediated by environmental selection)<sup>7</sup>. In microorganisms, phylogeographic studies have revealed that the genetic similarity within particular populations decreases with distance, which suggests a significant influence of dispersal limitation, leading to strong endemism<sup>3</sup>, and contradicts the hypothesis of microbial cosmopolitanism<sup>8</sup>. At the community level, biogeographical studies have indicated that diversity turnover is influenced both by geographical distance<sup>9</sup> and environmental heterogeneity (as classified by soil, vegetation and climatic characteristics)<sup>10,11</sup>. However, few in-depth studies have been carried out on the TAR of soil microorganisms, especially for soil bacteria, leading to a lack of concepts concerning the diversification processes.

TAR was first reported for plants in 1921 by Arrhenius<sup>12</sup> as a power-law relationship:

$$S_A = S_0 A^z, \quad (1)$$

where  $S_A$  is the number of taxa recorded in an area  $A$ ,  $S_0$  the initial number of taxa in the smallest sampling area and  $z$  is the rate at which new taxa are sampled as the sampling area is increased. TAR was extended to microorganisms less than a decade ago<sup>3,9,13</sup> by focusing on community similarity between sites:

$$\chi_d = \chi_D \left( \frac{d}{D} \right)^{-2z}, \quad (2)$$

where  $z$  is the turnover rate as in equation (1),  $\chi_d$  and  $\chi_D$  are the operational taxonomic unit ( $OTU_{bin}$ ) similarities between sites located  $d$  meters and  $D$  meters apart from each other ( $d^2$  and  $D^2$  would correspond to the area of the sampled locations), respectively. Nowadays, TAR is commonly used in this form in most microbial biogeographical studies<sup>3,9,14–16</sup> to assess microbial diversification and its potential relative dependency on “dispersal” and “habitat heterogeneity” (including habitat diversity and landscape configuration)<sup>4</sup>. Nevertheless, the TAR is still highly debated, particularly regarding its form, underlying hypotheses and factors affecting the relationship<sup>17–19</sup>. Rosindell *et al.*<sup>19</sup> highlighted that the TAR follows a power-law relationship under the hypothesis of infinite community size and for a steady  $z$ . As the soil microbial community is very large and diverse<sup>20</sup>, the relationship for these organisms exhibits power-law behaviour because the average abundance per microbial taxonomic unit is large according to Harte *et al.*<sup>21</sup> On the other hand, the constant  $z$

hypothesis across spatial scales, which assumes self-similarity as a probability rule<sup>13,18</sup> for the spatial distribution of taxa abundance, may not hold for soil microbes<sup>15,21,22</sup>. Therefore,  $z$  may be considered constant only at a given scale of investigation, not across spatial scales<sup>21</sup>, and such spatial variations must be specified for microbial communities.

Now, it is crucial to investigate whether spatial variations of soil bacterial communities observed on a wide scale comply with the Baas Becking postulate<sup>1</sup> involving solely environmental selection as related to habitat-related environmental heterogeneity, or if other ecological processes could be involved in these patterns (for example, dispersal in the context of Neutral Theory). In this context, at least two objectives might be reached, (i) determine the extent to which soil bacterial diversity increases with sampling area by applying TAR from local to wide scales<sup>6,23</sup>, and (ii) evaluate the link between biodiversity turnover, environmental heterogeneity and dispersal by confronting community and habitat diversity turnover rates.

In this study, we attained these objectives by conducting a wide-scale investigation on 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 of whole France (2,085 sites analysed)<sup>24</sup>. Variations in microbial community between soils were assessed by bacterial Automated Ribosomal RNA Intergenic Spacer Analysis (ARISA) directly on soil DNA<sup>25</sup>. The TAR was then applied to ARISA data by considering the slope ( $z_{bacteria}$ ) together with the initial similarity<sup>5</sup> ( $\chi_{16}$ ) at the grain size of the study (16 km) and the average similarity ( $\bar{\chi}$ ) as proposed by Morlon *et al.*<sup>4</sup> to complete the analysis of community spatial variations. In parallel, soil habitat, namely the “description of a physical place, at a particular scale of space and time, where an organism either actually or potentially lives”<sup>26</sup>, was characterized by including pedoclimatic, geomorphologic and land-use data. The TAR concept was then transposed to habitat, assuming that its similarity would decrease with increasing distance (habitat-area relationship (HAR)). For TAR and HAR, the dependence of the slopes and average similarities on spatial scales was evaluated on the scale of France, after which the study of ecological processes was focused on a circular neighbourhood 280 km in diameter. The slopes of the TAR and HAR, as well as the initial similarities and average similarities at this spatial scale, were then confronted and highlighted that dispersal limitation and environmental heterogeneity were not mutually exclusive to determine the bacterial diversity turnover.

## Results

**Soil-sampling network.** The French Soil Quality Monitoring Network represents 2,085 analysed sites, 2,172,570 pairwise comparisons with distances ranging from 16 km to *ca.* 1,100 km and covers more than 550,000 km<sup>2</sup>. Through this network, the enormous environmental heterogeneity of France was examined from a huge number of combinations of soil physico-chemical characteristics, geomorphologic characteristics, land uses and climatic conditions (Table 1 and Fig. 1).

The range of spatial structures for soil bacterial communities and soil habitat characteristics was estimated by computing experimental variograms for each axis of the principle component analysis (PCA) analyses of bacterial community data and habitat data. The results are presented in Table 2. The nugget effects were large for bacterial communities (nugget:sill ratio of 0.57 to 0.85) and smaller for soil habitat (nugget:sill ratio of 0.21 to 0.68). The effective variations for the different axes, both for habitat and soil bacterial community, ranged from 150 to 470 km and highlighted large spatial structures. The size of the spatial structures

**Table 1 | Summary statistics of habitat variables.**

	<b>n</b>	<b>Mean (<math>\pm</math> s.e.)</b>	<b>Range</b>	<b>CV (%)</b>
Clay ( $\text{g kg}^{-1}$ )	2,085	$245.2 \pm 2.9$	(2–819)	54
Silt ( $\text{g kg}^{-1}$ )	2,085	$406.2 \pm 3.8$	(2–819)	43
Sand ( $\text{g kg}^{-1}$ )	2,085	$348.6 \pm 5.1$	(7–986)	67
pH <sub>water</sub>	2,085	$6.4 \pm 0.03$	(3.7–9.2)	21
$C_{\text{org}}$ ( $\text{g kg}^{-1}$ )	2,085	$25.4 \pm 0.4$	(0.6–243.0)	80
Total N ( $\text{g kg}^{-1}$ )	2,085	$2.1 \pm 0.03$	(0–16)	73
$\text{CaCO}_3$ ( $\text{g kg}^{-1}$ )	2,085	$53.8 \pm 2.9$	(0–866)	250
C:N	2,085	$12.2 \pm 0.1$	(4.6–52.7)	37
Mean annual temperatures ( $^{\circ}\text{C}$ )	2,085	$11.6 \pm 0.03$	(0.2–16.2)	14
Sum of annual rainfall (mm)	2,085	$935.3 \pm 4.9$	(535.7–2158.2)	24
Annual evapotranspiration (mm)	2,085	$811.9 \pm 3.0$	(501.6–1560.3)	17
Altitude (m)	2,085	$326 \pm 8.8$	(0–2980)	124
Slope (%)	2,085	$6.6 \pm 0.2$	(0–103.2)	159
Orientation ( $^{\circ}$ )	2,085	$179.9 \pm 2.4$	(–1–359.5)	60
Land uses				
Cropplands	869			
Forests	567			
Vineyards	36			
Grasslands	523			
Others	90			

CV, coefficient of variation. Range: (minimum–maximum). Orientation corresponds to that of the largest slope in the grid cell. 'Others' for land uses represents particular land use such as orchards, wild lands, urban parks and specific ecosystems like upland meadows.

explaining the most the inertia of bacteria and habitat variations (PCA axis 1 in Table 2) ranged from 250 to 320 km.

**Distance decay of bacterial community and habitat.** The TAR measures the spatial turnover of community richness and allows the rate of community composition turnover (commonly referred to as  $z$ ) to be calculated. According to Green *et al.*<sup>9</sup>,  $z$  can be equivalently calculated from the increase of community richness with sampled area or from the decay of similarity in the community composition with increasing distance between sampling sites. Here the bacterial community composition of the soil samples was characterized by ARISA and the samples were sorted into OTUs.

Bacterial community similarity on the scale of France decreased significantly with geographical distance, indicating that the TAR was significant and demonstrating a spatially structured distribution of soil bacteria over a distance ranging from 16 to 1,100 km (Fig. 2a). At this scale, the estimated  $z_{\text{bacteria}}$  was  $0.006 (\pm 7 \times 10^{-4}, P < 0.001)$ .

The turnover rate of habitat ( $z_{\text{habitat}}$ ) was determined by transposing the TAR to habitat, assuming that the HAR was a power-law relationship (Fig. 2b). Similarly, a significant HAR was also found on the scale of France with an estimated  $z_{\text{habitat}}$  of  $0.023 (\pm 1 \times 10^{-5}, P < 0.001)$ .

**Influence of spatial scale on TAR and HAR.** As the parameters (initial similarity and turnover rate) of the TAR, and therefore those of HAR by transposition, might be affected by the spatial scale considered, the regression parameters of the TAR and HAR were compared for different spatial scales (0–70, 70–140, 140–280, 280–560 and 560–1120 km, Fig. 3). The  $z_{\text{bacteria}}$  ranged from 0.003 to 0.010, depending on the scale. The  $z_{\text{bacteria}}$  was not significantly different at the smallest (0–70 and 70–140 km) and the largest (560–1120 km) scales, but was significantly higher than at the intermediate scales (140–280 and 280–560 km). Conversely, the  $z_{\text{habitat}}$  presented higher values than the  $z_{\text{bacteria}}$  (from 0.018 to 0.055) and was positively correlated with spatial scale. Interestingly,  $z_{\text{habitat}}$  increased moderately and regularly up to the

280–560 km scale, but increased very drastically at scale larger than 580 km.

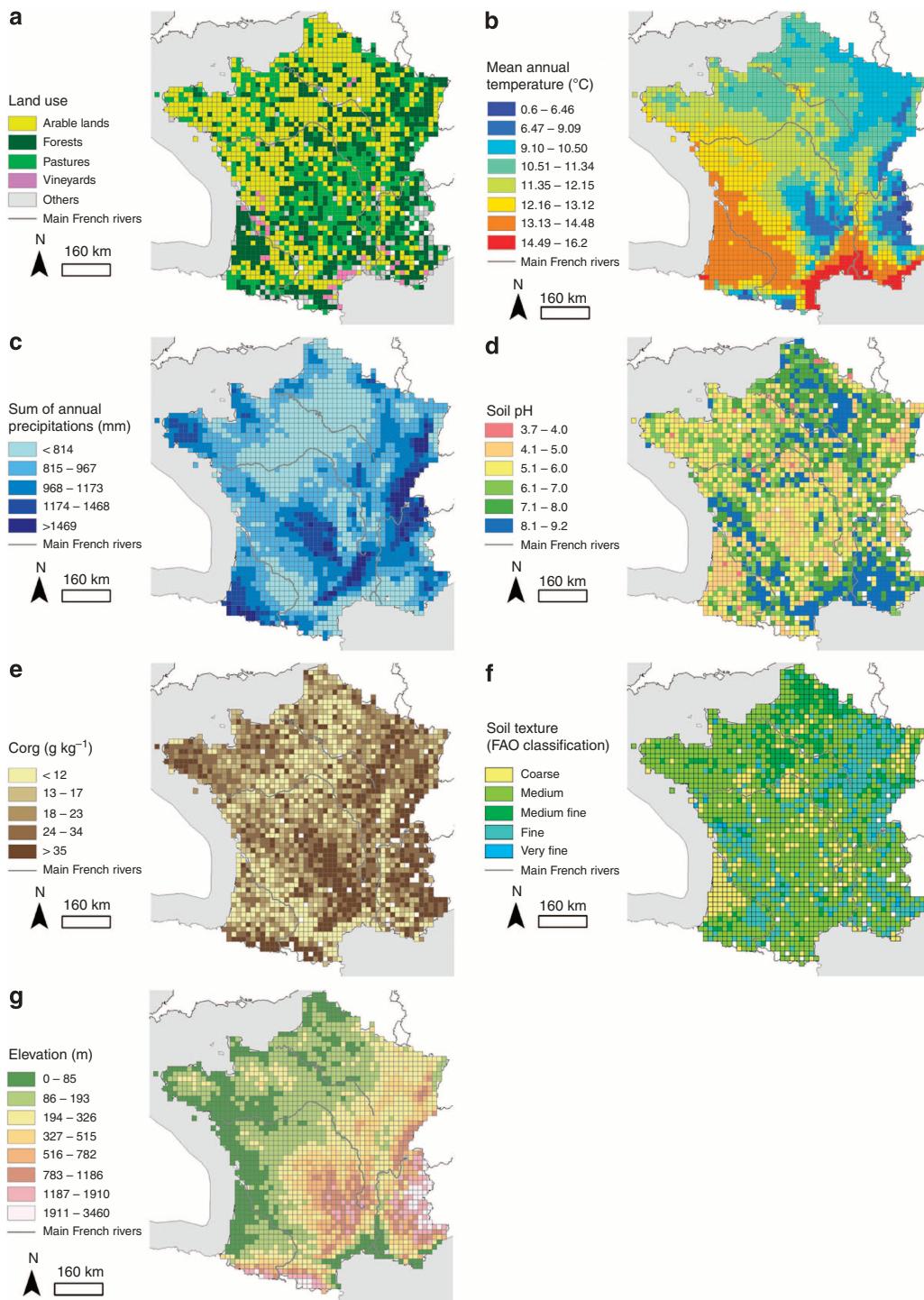
The average similarities of bacterial communities and habitat ranged from 0.62 and 0.64 to 0.71 and 0.81, respectively. The ranges of variations were small, even if significant differences were observed between every spatial scale with a decreasing trend from small to large scales. The initial similarities of bacterial community ( $\chi_{16}$ ) or habitat ( $E_{d16}$ ) were defined at the grain size of the study (16 km) and could not be calculated over multiple spatial scales. Nevertheless, this metric was compared with the average similarity at different scales. Here,  $\chi_{16}$  and  $E_{d16}$  were 0.65 and 0.84, respectively, representing higher values than the average similarities calculated at different scales.

**Relationship between initial or average similarities.** Here the objective was to determine if bacterial similarity was determined by the similarity of habitat at a local spatial scale (16 km) and at the patch scale, that is, if similar habitats house similar bacterial communities. The confrontation of initial similarities ( $\chi_{16}$  and  $E_{d16}$ , Fig. 4a) exhibited a relationship modelled with a linear regression type II model. The associated permutation test highlighted the significance of this relationship (1,000 permutations,  $r^2 = 0.123, P < 0.001$ ). In this relationship,  $\chi_{16}$  decreased with increasing  $E_{d16}$ , but the range of variations of  $\chi_{16}$  and  $E_{d16}$  remained small.

The average similarity of bacterial community structure was significantly correlated to the average habitat similarity ( $r^2 = 0.036, P < 0.05$ , 1,000 permutations Fig. 4b), but the range of variations of  $\bar{\chi}$  and  $\bar{E}_d$  remained small.

In addition,  $\chi_{16}$  and  $\bar{\chi}$ , and  $E_{d16}$  and  $\bar{E}_d$  were also correlated ( $r = 0.982$  and  $0.806$ , respectively).

**TAR–HAR relationship.** The slopes of the TAR and HAR were confronted to evaluate if the turnover rate of soil bacterial community was determined by the turnover rate of soil habitat. To achieve this goal, the statistical confrontation of  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$  was performed by systematically defining TAR and HAR in a neighbourhood corresponding to a circular sliding window of



**Figure 1 | Environmental variability across France.** (a) Map of land-uses according to the Corine land cover classification; (b, c) maps of climatic conditions: mean annual temperature and sum of annual precipitations, respectively; (d-f) Maps of soil physico-chemical characteristics: pH, organic carbon content and soil texture according to the Food and Agriculture Organization (FAO) classification; (g) map of elevation (m).

280 km diameter, centred on each of the 2,085 RMQS sites. The window size was deduced from the size of spatial structures of bacteria and habitat identified by the experimental variograms on the scale of France (Table 2)<sup>27</sup>.

The range of variations in  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$  could thus be determined and showed that  $z_{\text{bacteria}}$  ranged from  $-0.003$  to  $+0.021$  (mean = 0.008, median = 0.009), whereas  $z_{\text{habitat}}$  ranged from  $-0.002$  to 0.131. For  $z_{\text{bacteria}}$ , the negative turnover rates

(<1% of total estimates) were observed in some regions where distant sites were more similar to those in the centre, than to intermediate sites. Although in 62% of the regions defined by the sliding windows, the TAR ranged from 0.002 to 0.01 and was close to that observed for France as a whole, in 36% of the regions, the TAR was significantly higher and ranged from 0.01 to 0.02. In other respects, the value of  $z_{\text{habitat}}$  calculated for France (0.023) was lower than the mean (0.040) or median (0.035)

obtained for the 2,085 regions. In some regions ( $<1\%$ ),  $z_{\text{habitat}}$  was negative because distant sites showed greater similarity to those in the centre than to intermediate sites.

Statistical confrontation of the 2,085  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$  estimates revealed a linear relationship (Fig. 5). The significance of the relationship was tested through a linear regression type II model<sup>28</sup>, as both estimated turnover rates were associated with errors. A permutation test (random association of  $z_{\text{habitat}}$  and  $z_{\text{bacteria}}$  values, 1,000 permutations)<sup>29</sup> showed that the model and the regression parameters were highly significant ( $r^2 = 0.667$ ,  $P < 0.001$  and Fig. 5).

**Influence of landscape configuration.** Habitat turnover rate and initial similarity are determined by habitat diversity and its spatial configuration (that is, landscape configuration). In the previous section, the permutation test demonstrated that habitat diversity significantly influenced the relationship between habitat and bacterial community turnover and initial similarities. Therefore, in this section the objective was to evaluate the importance of landscape configuration on these relationships. To achieve this goal, a permutation test was performed in which the spatial organization of habitat was reconfigured without modifying the bacterial community variability, habitat variability or the selection pressure of habitat on the bacterial community (pairs of habitat and bacterial community composition data unmodified). The

modification of landscape configuration significantly altered the correlation (Fig. 6a, b) between  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$  and changed the sign of the correlation between  $\chi_{16}$  and  $E_{d16}$  in the random case. This might be related to the range of variations of  $\chi_{16}$ , which was smaller than that of  $E_{d16}$ . Similarly, the intercepts of the relationships  $\beta_z$  and  $\beta_\chi$  were significantly reduced by landscape reconfiguration: 0.004 versus 0 and 0.820 versus 0.558, respectively (Fig. 6c, d). The slope of the relationship between the initial similarities ( $\alpha_\chi$ ) was also affected by landscape reconfiguration ( $-0.221$  versus  $0.108$  in the random case Fig. 6f), but this was not the case for the slope  $\alpha_z$  that harboured very similar values for the observed and the random cases (0.1350 and 0.1231, respectively, Fig. 6e).

## Discussion

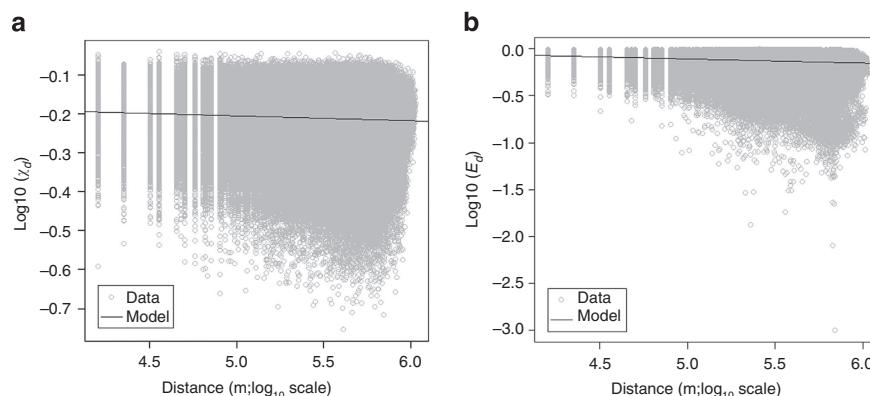
In this study, a significant turnover of soil bacterial diversity was demonstrated on a broad scale. This is an important and relevant finding, as several aspects of bacterial physiology and ecology (ubiquity, horizontal gene transfer, high ability to colonize extreme environments, redundancy of ecological attributes)<sup>20</sup> would be expected to prevent a significant TAR for these organisms. Significant TAR have already been reported for indigenous bacteria from various ecosystems, such as water-filled tree holes<sup>30</sup> or salt marshes<sup>14</sup>, which suggests that TAR is a universal relationship for all living organisms and all environments.

The estimated  $z_{\text{bacteria}}$  was  $0.006 (\pm 7 \times 10^{-4}, P < 0.001)$ , that is, within the lower range of turnover rates for microbes ( $0.002 < z < 0.26$ ), as compiled by Woodcock *et al.*<sup>31</sup> This is in agreement with the general trend observed for microbes, the values usually being well below than those of the macroorganisms ( $z_{\text{macroorganisms}} \geq 0.1$ )<sup>14</sup>, even if a higher  $z_{\text{bacteria}}$  (0.26) has been reported for small, discrete and highly heterogeneous ecosystems<sup>30</sup> known as hosting insular communities with a high community turnover<sup>32</sup>. The significant but low turnover observed on a wide scale in our study might partly be ascribed to the high average abundance per taxonomic unit, which induces low turnover rates<sup>21</sup>, and to technical limitations, particularly the low taxonomic resolution of DNA fingerprinting. Indeed, DNA fingerprints correspond to dominant bands, each including multiple species. This results in an aggregation of species composition variations within the bands and may contribute to the low turnover rate observed<sup>4</sup>. In addition, DNA fingerprinting precludes consideration of the accumulation of new minor species

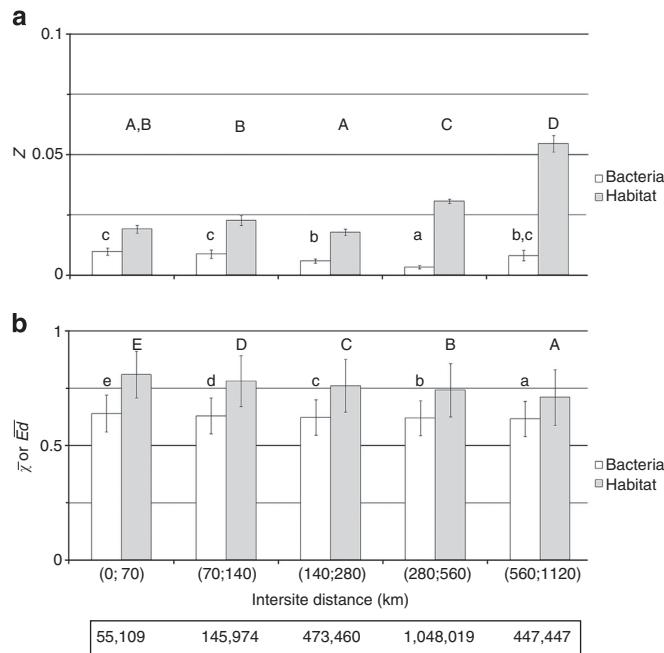
**Table 2 | Fitted parameters of the variograms of bacterial genetic structure and soil habitat.**

Data	PCA axis number	Inertia (%)	S=nugget:sill ratio	Effective range (km)
Bacterial community structure	1	22.6	0.57	249.6
	2	10.7	0.79	166.4
	3	6.9	0.85	416.1
	4	6.4	0.78	249.6
	5	5.7	0.78	249.6
Soil habitat	1	21.4	0.21	323.4
	2	17.7	0.46	154.7
	3	11.2	0.43	421.8
	4	7.3	0.68	469.5
	5	6.5	0.67	295.3

PCA, principle component analysis. The different parameters were computed for each axis of the PCAs performed on genetic and environmental data, respectively.



**Figure 2 | TAR and HAR relationships at the scale of France.** (a) TAR, (b) HAR, grey circles: logarithm of genetic and habitat similarity values for each paired comparison of sites confronted with the logarithm of between-sites geographical distance (m); black line: weighted linear regression model for the distance-decay relationships. (a) Significant model ( $P < 0.001$ ):  $\log_{10}(\text{genetic similarity}) = -0.0117 \cdot \log_{10}(\text{geographical distance}) - 0.1464$ . (b) Significant model ( $P < 0.001$ ):  $\log_{10}(\text{habitat similarity}) = -0.0454 \cdot \log_{10}(\text{geographical distance}) + 0.1174$ .

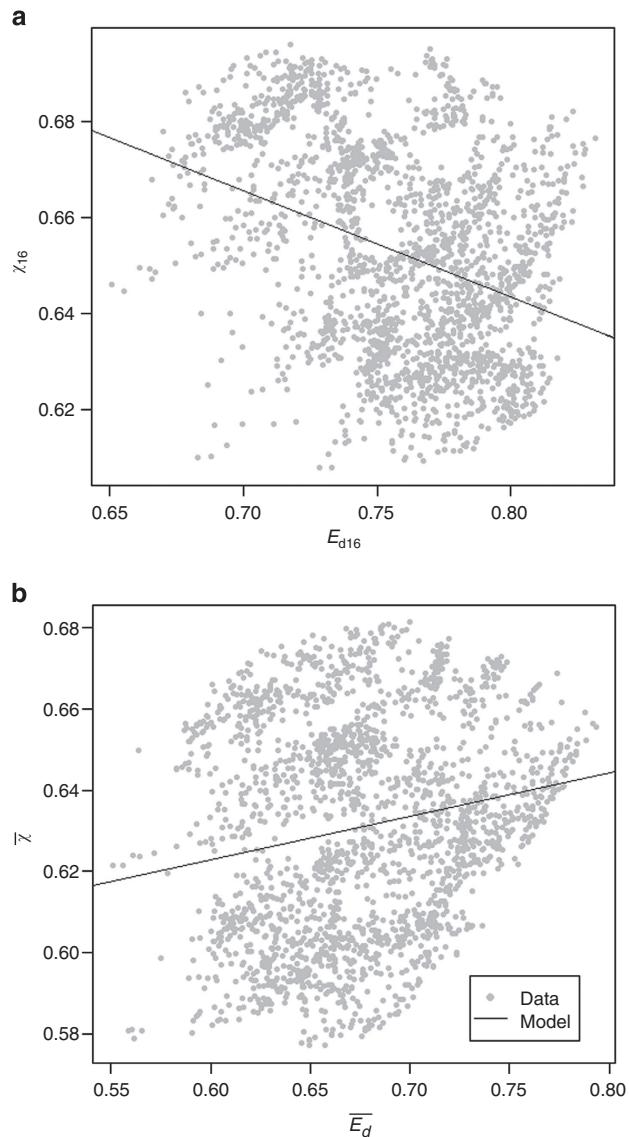


**Figure 3 | Scale effects on TAR and HAR parameters.** Variations of  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$  (a), and of the average similarity of bacterial community ( $\bar{\chi}$ ) or habitat ( $\bar{E}_d$ ) (b) at different spatial scales. Error bars correspond to 95% intervals for regression parameters ( $z$ ) and to s.d. for average similarity of bacterial community or habitat. The result of the statistical comparisons of  $z$  or average similarities between spatial scales is indicated through capital letters for habitat and through lower case letters for bacteria. The box at the bottom of the figure indicates the number of pairwise comparisons per spatial scale.

with increasing distance, which may represent a large proportion of the total soil microbial species richness<sup>33</sup> and determine the slope of TAR<sup>14,31</sup>. The low  $z_{\text{bacteria}}$  might also be related to the grain size of our sampling design (16 km × 16 km: 256 km<sup>2</sup>), which did not consider scales less than landscape, and smoothed significant local variations in soil microbial community composition that might potentially affect TAR<sup>34</sup>.

In parallel, soil habitat was characterized by a multivariate analysis including pedoclimatic, geomorphologic and land-use data, and the HAR was calculated by transposing the TAR concept. On the scale of France,  $z_{\text{habitat}}$  was  $0.023 (\pm 1 \times 10^{-5}, P < 0.001)$ . This value can only be considered in relation to  $z_{\text{bacteria}}$ , as no other means of comparison is available in the literature. The  $z_{\text{habitat}}$  was higher than the  $z_{\text{bacteria}}$ , which might be related to the lower average abundance per habitat than per bacterial OTU<sub>bin</sub>.

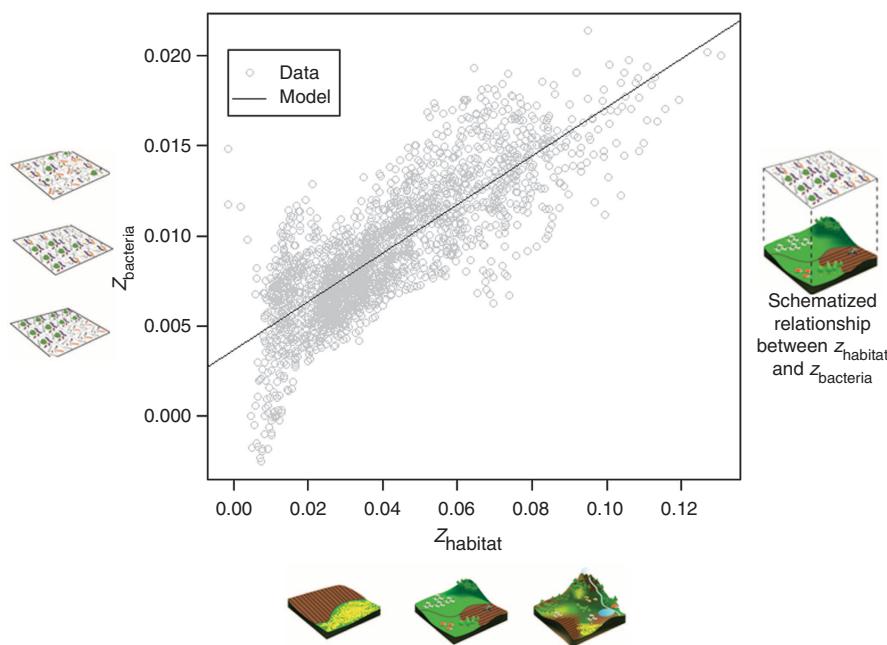
As described in previous studies, TAR is sensitive to the spatial scale of investigation<sup>21</sup>. The sampling design applied in our study allowed us to test this hypothesis on both TAR and HAR, and to consider multiple spatial scales (distance range between site: [0; 70], [70–140], [140; 280], [280; 560] and [560; 1100] km). The significant effect of spatial scale on the average similarities of soil bacterial community and habitat and on the turnover rates was demonstrated. The average similarities decreased significantly with increasing spatial scale and were always lower than the corresponding initial similarity (0.65 and 0.84 for  $\chi_{16}$  and  $E_{d16}$ , respectively). This is in agreement with the hypothesis that neighbouring sites are more similar than distant sites<sup>2,15</sup>. The significant decrease of  $z_{\text{bacteria}}$  with increasing spatial scales might be explained by the higher average abundance per OTU with increasing scales<sup>21,22</sup>. Similarly, the significant increase of  $z_{\text{habitat}}$



**Figure 4 | Relationship between bacterial communities and habitat similarities.** (a) Initial similarities at 16 km; and (b) mean similarity of habitat and bacterial community estimated for each RMQS site considering every site within a circular neighbourhood with a 280 km in diameter. Black line: best significant model fitted to the data ( $P < 0.001, 1,000$  permutations): (a)  $\log_{10}(\chi_{16}) = -0.2210 * \log_{10}(E_{d16}) + 0.8204$ , (b)  $\bar{\chi} = 0.1071 * \bar{E}_d + 0.5586$ .

with spatial scale might result from decreasing average abundance per habitat, as the number of observations per habitat increases equally or even more slowly than habitat richness.

Altogether, the significant variations of  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$  according to scale might reflect variations in the relative importance of diversification processes<sup>4</sup>. Morlon *et al.*<sup>4</sup> demonstrated the relevance of distance–decay relationships for testing spatial ecology theories. We therefore compared three metrics relating bacterial community and habitat, namely turnover rates, initial and average similarities, to identify the processes involved in the spatial distribution of soil bacterial communities. These comparisons were performed by computing and compiling data from the 2,085 circular neighbourhoods (280 km in diameter as determined from experimental variograms, see Table 2). At this spatial scale, the main



**Figure 5 | Relating turnover rates of soil bacterial community structure and habitat.** Grey circles represent the turnover rates calculated for each RMQS site considering every site within a distance of 140 km in a circular window. The black line represents the linear model type II fitted to the data:  $z_{\text{bacteria}} = 0.1350 * z_{\text{habitat}} + 0.0036$ , ( $r^2 = 0.65$ ,  $P < 0.001$ ). Representation of diversity and spatial fragmentation of soil bacterial community assemblage and of habitats are schematized by the y axis and under the x axis, respectively.

processes to consider were (i) environmental heterogeneity leading to adaptation of populations to a particular habitat, the larger areas exhibiting a higher renewal of habitat variation<sup>2,30</sup>, that is, the “environmental selection hypothesis”, and (ii) “dispersal limitations” resulting from the balance between species extinction and colonization, which is significantly influenced by the size of the sampling area. If dispersal is infinite, community composition is fully determined by environmental selection, bearing out the postulate of Baas Becking<sup>1</sup>: “Everything is everywhere, but, the environment selects”. Under this postulate, the turnover rate of the soil bacterial community would be entirely proportional to the turnover of the soil environment and would, on average, be zero in completely homogeneous soil environments ( $z_{\text{habitat}} = 0$ ). Otherwise, if the dispersal of soil bacteria is limited, significant turnover rates of bacterial community would be observed in homogeneous habitats according to the Neutral Theory<sup>19,35</sup>. Under infinite dispersal, the same interpretation would hold for the relationships between the initial or average similarities of soil bacterial community and habitat, which should be positive and reach the maximum in fully homogeneous habitat.

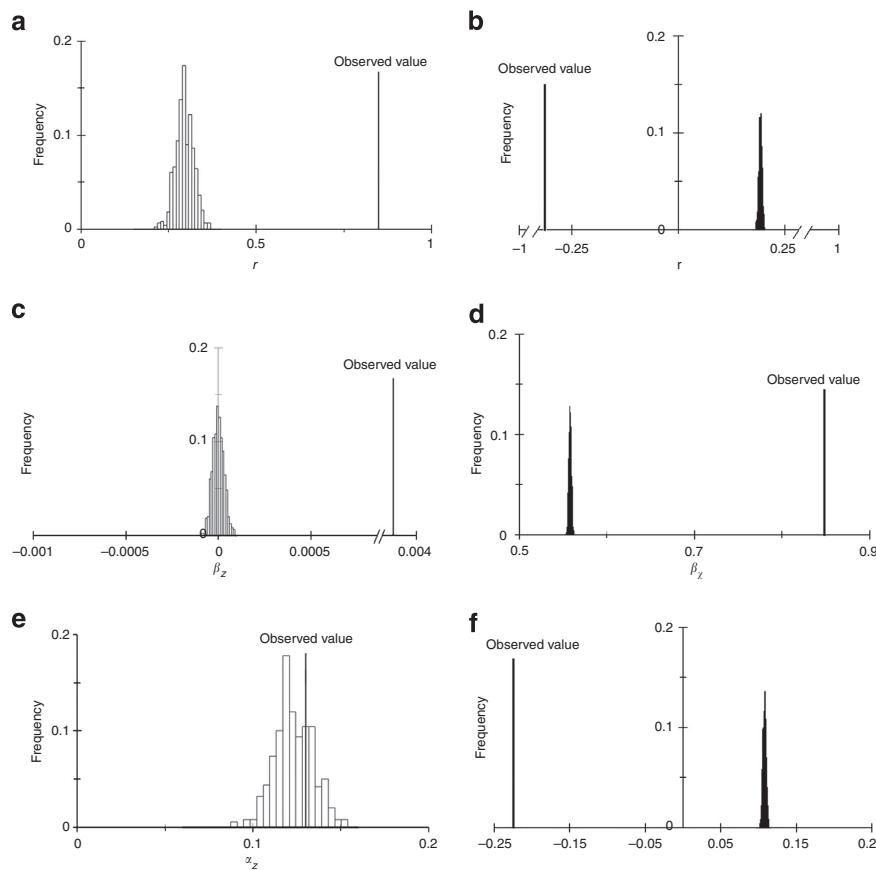
Here the average habitat similarity and average bacterial community similarity were significantly and positively correlated at the scale of the circular neighbourhoods, which was in agreement with the environmental selection hypothesis<sup>2,16,36</sup>. Nevertheless, the initial similarities of habitat and soil bacterial community were weakly but negatively correlated, suggesting that variations in community composition occur at small scales, even in homogeneous habitats.

The significant linear regression between  $z_{\text{habitat}}$  and  $z_{\text{bacteria}}$  demonstrated a positive relationship between soil bacterial community turnover and habitat turnover on a regional scale (Fig. 5), thereby supporting the environmental selection hypothesis for soil microorganisms<sup>3,7</sup>. It is interesting to note that  $z_{\text{bacteria}}$  values may attain those reported for larger organisms (*ca.* 0.02), even in large areas of contiguous habitats, suggesting that

communities of both bacteria and macroorganisms are structured by analogous processes<sup>31</sup>. Recent microbial biogeography studies have revealed that the main environmental filters shaping spatial microbial diversity distribution are soil physico-chemical characteristics, land use and plant cover, whereas climatic and geomorphologic filters are less important<sup>9,11,15,37</sup>. Similar filters were also reported to explain soil bacterial community distribution in contrasted ecological regions at the scale of France<sup>27,34</sup>.

Nevertheless, the estimated  $z_{\text{bacteria}}$  in the case of a homogeneous habitat ( $z_{\text{habitat}} = 0$ ) was not equal to 0, suggesting that bacterial communities might have a limited dispersion despite the huge amounts of bacteria ( $>10^{15}$  cells per year) transported through the atmosphere between continents<sup>38</sup>. This conclusion for the soil bacterial community as a whole is in agreement with phylogeographic studies focusing on particular soil bacterial groups<sup>15,16,39</sup> and with studies showing a distance effect on soil bacterial community structure<sup>2</sup>.

We then determined the relative dependency of bacterial community turnover on dispersal limitation by testing the effect of habitat spatial configuration on the relationships between  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$ , and between the initial similarities ( $\chi_{16}$  and  $E_{16}$ ) without modifying the environmental selection process, that is, the pairs of soil habitat characteristics and associated bacterial community composition data. By shuffling habitat spatial configuration, we altered the relationships between turnover rates and between initial similarities of bacterial community and habitat, respectively (Fig. 6). The versatility of the correlation (in terms of sign and significance) between the initial similarities confirmed the difficulty of drawing conclusions at the grain size of this study. On the other hand, the relationship between turnover rates was significantly altered for the correlation and intercept, but not for the slope. This supported the hypothesis that the slope of the relationship between turnover rates is determined by the selection pressure of the soil habitat on the indigenous bacterial community. It also suggested that the intercept of the relationship (bacterial diversity turnover rate in



**Figure 6 | Effect of landscape on microbial and habitat turnover rates or initial similarities.** (a) Distribution of the correlation coefficient ( $r$ ) of the relationship between turnover rates; (b) distribution of the correlation coefficient ( $r$ ) of the relationship between initial similarities; (c) distribution of the intercept ( $\beta_z$ ) of the relationship between turnover rates; (d) distribution of the intercept ( $\beta_z$ ) of the relationship between initial similarities; (e) distribution of the slope ( $\alpha_z$ ) of the relationship between turnover rates; (f) distribution of the slope ( $\alpha_z$ ) of the relationship between initial similarities. Each distribution is based on 500 permutations in each sliding window.

homogeneous habitats) is related to dispersal limitations mediated by the spatial configuration of the habitat. This statement would be in favour to the role of population aggregation in determining the turnover rate of bacterial diversity<sup>4</sup>.

Our study negates the hypothesis that the soil microbial community is a “black box” with no spatial structure and exhibiting a homogeneous distribution from local to large scales. As observed in macroorganisms, spatial patterns in soil bacterial communities are also significantly governed by environmental heterogeneity and dispersal, and the two processes are not mutually exclusive<sup>2,7</sup>. Moreover, the environmental filters impacting the turnover of soil microbial diversity as a whole need to be deciphered and ranked to better understand the sustainability of the soil ecosystem services provided by the corresponding biodiversity. This represents a major challenge in microbial ecology, but should provide the knowledge required for sustainable soil management and implementation of the corresponding protection policies on a wide scale<sup>40</sup>.

## Methods

**Soil-sampling strategy.** The soil samples were provided by the platform GenoSol ([http://www.dijon.inra.fr/plateforme\\_genosol](http://www.dijon.inra.fr/plateforme_genosol)) from the soil storage facility of the RMQS (French Monitoring Network for Soil Quality), which is a soil-sampling network based on a 16 km × 16 km systematic grid covering the whole of France<sup>31</sup>. Each of the 2,195 monitoring sites has been precisely geopositioned, and the soil profile, site environment, climatic factors, vegetation and land use have been accurately described. In the centre of each 16 km × 16 km square, 25 individual core samples were taken from the topsoil (0–30 cm) within a 20 m × 20 m area, using an unaligned sampling design. The core

samples were then bulked to obtain a composite sample for each site. The soil samples were gently air-dried, sieved to 2 mm and stored at –40 °C before DNA analysis.

Physical and chemical analyses (listed in Table 1 and Fig. 1) are available for 2,131 soils and were performed by the Soil Analysis Laboratory of INRA (Arras, France, <http://www.lille.inra.fr/las>). Available climatic data were monthly rain, Potential EvapoTranspiration and temperature at each node of a 12 km × 12 km grid, averaged for the 1992–2004 period. These climatic data were obtained by interpolating observational data using the SAFRAN model<sup>42</sup>. The RMQS site-specific data were linked to the climatic data by finding the closest node from the 12 km × 12 km climatic grid for each RMQS site. Land cover was recorded according to the CORINE Land Cover classification at level 1 (<http://www.ifen.fr>), which consists of a rough descriptive classification of land use into five classes: arable lands, forests, pastures, vineyards and other land uses. All these data were available for 2,085 soils in the INRA InfoSol DONESOL database ([http://www.gissol.fr/programme/rmq/RMQS\\_manuel\\_31032006.pdf](http://www.gissol.fr/programme/rmq/RMQS_manuel_31032006.pdf); Table 1 and Fig. 1).

**Characterization of habitat.** Soil habitat was characterized for each site by its coordinates in a Hill and Smith multivariate analysis<sup>43</sup> using the dudi.hillsmith function in the ade4 package<sup>44</sup> in R, which allows mixing quantitative (geomorphological data: altitude, slope and orientation; climatic data: mean annual temperature, annual rainfall and annual potential evapotranspiration; soil physico-chemical data: clay, silt and sand contents, pH<sub>water</sub>, organic C content, total N content, CaCO<sub>3</sub> content and C:N ratio) and qualitative variables (land use, five classes; Table 1 and Fig. 1). Quantitative variables were centred and scaled, and the qualitative variables were converted into binary variables. All the quantitative variables had a weight of 1 in the analysis and the binary variables had a weight equal to 1/ $n$ ,  $n$  being the number of classes observed in the qualitative variables they were derived from.

According to the decrease of inertia per axis, five axes representing 63.9% of the variance in the environmental data were retained to characterize the habitat of each

study site. The coordinates of the sites in the Hill and Smith multivariate analysis were used in the following and referred to as habitat characteristics.

**Molecular characterization of bacterial community structure.** DNA was extracted from 2,085 soils from the RMQS grid, using a single procedure optimized by Ranjard *et al.*<sup>25</sup>, which is reliable and robust for the routine analyses of several hundreds of different soils. Soil bacterial community structure was genotyped directly from soil DNA extracts using a bacterial ARISA, fingerprinting approach optimized for medium throughput in the GenoSol platform.

Within the ARISA method, the bacterial intergenic spacers were amplified with the primer set: S-D-Bact-1522-b-S-20/ L-D-Bact-132-a-A-18, with 50 ng of DNA as template, the PCR conditions being described by Ranjard *et al.*<sup>25</sup>. The S-D-Bact-1522-b-S-20 primer was labelled at its 5' end with the IRD800 dye fluorochrome (MWG SA Biotech, Ebersberg, Deutschland) to allow detection of the PCR fragments by the LiCor DNA sequencer system (ScienceTec, Les Ulis, France). PCR fragments were resolved on 3.7% polyacrylamide gels run under denaturing conditions for 15 h at 3,000 V/60 W on a LiCor DNA sequencer (ScienceTec). The data were analyzed using the 1D-Scan software (ScienceTec). This software converts fluorescence data into electrophoregrams where the peaks represent the PCR fragments. The height of the peaks was calculated in conjunction with the median filter option and the Gaussian integration in 1D-Scan, and represented the relative proportions of fragments in the total products. Lengths (in base pairs) were calculated using a size standard with bands ranging from 200 to 1,206 bp.

Data obtained from the 1D-Scan software (ScienceTec) were converted into a table summarizing band presence (that is, peaks) and intensity (that is, height or area of peak) using the PrepRISA program<sup>25</sup>. As described in a previous study, 100 peaks, 2 bp resolution and Gaussian peak area were used to provide a robust analysis of bacterial communities<sup>25</sup>. The resulting bacterial-ARISA data matrix (bacterial communities as rows and bands as columns), takes into account the presence/absence and relative intensity of bands.

**Bacterial community and habitat experimental variograms.** We used the classical geostatistics approach<sup>45</sup> to characterize the spatial process of bacterial genetic structure and soil habitat. Exponential variograms have been fitted using weighted least squares on the different experimental variograms computed for each axis of the PCAs performed on genetic and environmental data. In this method, the weights of each lag are the experimental variogram of the lag divided by the number of observation pairs within a bin centred on the lag. To characterize the spatial process, we have calculated first the nugget:sill ratio, which represents the non-explanatory part of the spatial model. Second, we retained also the effective range, which is the distance data that the semivariance value achieves 95% of its maximum.

**TAR and HAR computing.** The turnover rate of bacterial community composition ( $z_{\text{bacteria}}$ ) was determined by converting molecular fingerprint data into presence/absence data and computing between-sites similarity using the `dsvis` function (Sørensen index) of the `labdsv` package<sup>46</sup> in the R software<sup>47</sup>. Then,  $z_{\text{bacteria}}$  was estimated from the slope of the decay relationship between similarity and distance according to the formula:

$$\log_{10}(\chi_d) = (-2z) * \log_{10}(d) + b \quad (3)$$

where  $\chi_d$  is the observed Sørensen's similarity between two soil samples (number of common OTUs divided by the average number of OTUs in the two samples), which are  $d$  meters apart from each other,  $b$  is the intercept of the linear relationship and  $z$  is the turnover rate of the community composition (referred to as  $z_{\text{bacteria}}$ ). Regression parameters ( $-2z$ ) and  $b$  were estimated by weighted linear regression using the ordinary least squares method. The weight of each similarity value corresponded to the number of data in the respective distance class. This approach was applied to avoid bias related to large distance classes, including few but very dissimilar sites. The significance threshold was set at  $\alpha = 0.05$ .

The turnover rate of habitat ( $z_{\text{habitat}}$ ) was determined by transposing the TAR to habitat, assuming that the HAR was a power-law relationship. For this purpose, habitat characteristics were used to compute between sites similarity ( $E_d$ ) as follows:

$$E_d = \left( 1 - \frac{Euc_d}{Euc_{\max}} \right) + 0.001 \quad (4)$$

with  $Euc_d$  the Euclidean distance (`dist` function, R software) between two sites that are  $d$  meters apart from each other, and  $Euc_{\max}$  is the maximum Euclidean distance in the distance matrix; 0.001 was added to account for zero similarity between sites<sup>29</sup>. Then,  $z_{\text{habitat}}$  was estimated from the slope of the decay relationship between habitat similarity and distance according to equation (3). Regression parameters were calculated by adopting the same methodology used to calculate the parameters for the TAR.

**Average and initial similarities.** The average similarity of soil habitat ( $\bar{E}_d$ ) or of bacterial community composition ( $\bar{\chi}$ ) was determined as the mean of the similarity matrix computed for soil habitat or bacterial community as described above.

The initial similarities of soil habitat ( $E_{d0}$ ) or of bacterial community composition ( $\chi_0$ ) were computed as the mean similarity between sites for a distance between sites of 16 km.

**Relating habitat and bacterial similarities or turnovers.** To evaluate these relationships, a neighbourhood statistics approach was used allowing computation of  $z_{\text{bacteria}}$ ,  $z_{\text{habitat}}$ ,  $E_{d16}$ ,  $\chi_{16}$ ,  $\bar{E}_d$  and  $\bar{\chi}$  for each site in the RMQS on a circular neighbourhood 280 km in diameter. For that purpose, the same methods as those presented in the sections devoted to computing the TAR, the HAR, the average and the initial similarities of soil habitat and bacterial community composition were applied.

The size and shape of the neighbourhood was in agreement with previous spatial analyses of habitat and genetic data that demonstrated significant spatial structures<sup>27</sup> when the radius was at least 140 km (Table 2). As all the parameters are estimated with errors, the relationships were best modelled by a type II linear regression model (ordinary least squares method). This was done using the 'lmodel2' function of the 'lmodel2' package<sup>28</sup> in the R software. A permutation test (1,000 permutations) was used to test the significance of the relationship and its parameters<sup>28</sup>. The slope and the intercept were referred as follows:  $\alpha_Z$ ,  $\beta_Z$ ,  $\alpha_{\bar{Z}}$ ,  $\beta_{\bar{Z}}$  for the relationship between  $z_{\text{bacteria}}$  and  $z_{\text{habitat}}$ ,  $\chi_{16}$  and  $E_{d16}$ ,  $\bar{\chi}$  and  $\bar{E}_d$ , respectively.

**Influence of landscape configuration.** To evaluate the influence of landscape on the relationship between habitat and bacterial community turnover or between initial similarities, a permutation test (500 permutations) was performed. In this test, the neighbourhood statistics approach described above was used, but the geographic coordinates of the sites in the neighbourhood region (280 km diameter) were shuffled without modifying the pairs of soil habitat characteristics and bacterial community composition data. This allowed creating randomly configured regions in which bacterial community variability (average similarity and variance), habitat variability (average similarity and variance) and selection pressure of habitat on bacterial community were kept unchanged (pairs of habitat and bacterial community composition data unmodified). As a consequence, only turnover rates and initial similarities were affected, but not the average similarities. Then, the relationships between turnover rates or between initial similarities were computed for each permutation. As a result, the distributions of the Pearson's correlation coefficient and of the regression parameters ( $\alpha_Z$ ,  $\beta_Z$ ,  $\alpha_{\bar{Z}}$ ,  $\beta_{\bar{Z}}$ ) were considered for each relationship. Then, the probability of observed value were derived from these distributions and confronted to the 5% significance threshold.

## References

1. Baas Becking, L.G.M. *Geobiologie Of Inleiding Tot De Milieukunde* (The Hague, The Netherlands, 1934).
2. Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B.H. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506 (2012).
3. Martiny, J.B.H. *et al.* Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
4. Morlon, H. *et al.* A general framework for the distance-decay of similarity in ecological communities. *Ecol. Lett.* **11**, 904–917 (2008).
5. Soininen, J., McDonald, R. & Hillebrand, H. The distance decay of similarity in ecological communities. *Ecography* **30**, 3–12 (2007).
6. Gleason, A.H. On the relationship between species and area. *Ecology* **3**, 158–162 (1922).
7. Kallimanis, A.S. *et al.* How does habitat diversity affect the species-area relationship? *Glob. Ecol. Biogeogr.* **17**, 532–538 (2008).
8. Finlay, B.J. & Clarke, K.J. Ubiquitous dispersal of microbial species. *Nature* **400**, 828 (1999).
9. Green, J.L. *et al.* Spatial scaling of microbial eukaryote diversity. *Nature* **432**, 747–750 (2004).
10. Angel, R., Soares, M.I.M., Ungar, E.D. & Gillor, O. Biogeography of soil archaea and bacteria along a steep precipitation gradient. *ISME J.* **4**, 553–563 (2010).
11. Fierer, N. & Jackson, R.B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **103**, 626–631 (2006).
12. Arrhenius, O. Species and area. *J. Ecol.* **9**, 95–99 (1921).
13. Harte, J., Kinzig, A. & Green, J. Self-similarity in the distribution and abundance of species. *Science* **284**, 334–336 (1999).
14. Horner-Devine, M.C., Lage, M., Hughes, J.B. & Bohannan, B.J.M. A taxa-area relationship for bacteria. *Nature* **432**, 750–753 (2004).
15. Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D. & Horner-Devine, M.C. Drivers of bacterial  $\beta$ -diversity depend on spatial scale. *Proc. Natl. Acad. Sci. USA* **108**, 7850–7854 (2011).
16. Ramette, A. & Tiedje, J.M. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc. Natl. Acad. Sci. USA* **104**, 2761–2766 (2007).
17. Maddux, R.D. Self-similarity and the species-area relationship. *Am. Nat.* **163**, 616–626 (2004).

18. Ostling, A., Harte, J., Green, J.L. & Kinzig, A.P. Self-similarity, the power law form of the species-area relationship, and a probability rule: a reply to Maddux. *Am. Nat.* **163**, 627–633 (2004).
19. Rosindell, J., Hubbell, S.P. & Etienne, R.S. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol. Evol.* **26**, 340–348 (2011).
20. Maron, P.-A., Mougel, C. & Ranjard, L. Soil microbial diversity: methodological strategy, spatial overview and functional interest. *C. R. Biol.* **334**, 403–411 (2011).
21. Harte, J., Smith, A.B. & Storch, D. Biodiversity scales from plots to biomes with a universal species-area curve. *Ecol. Lett.* **12**, 789–797 (2009).
22. Drakare, S., Lennon, J.J. & Hillebrand, H. The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol. Lett.* **9**, 215–227 (2006).
23. Green, J. & Bohannan, B.J.M. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**, 501–507 (2006).
24. Ranjard, L. *et al.* Biogeography of soil microbial communities: a review and a description of the ongoing french national initiative. *Agron. Sustain. Dev* **30**, 359–365 (2010).
25. Ranjard, L. *et al.* Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. *Environ. Microbiol.* **5**, 1111–1120 (2003).
26. Kearney, M. Habitat, environment and niche: what are we modelling? *Oikos* **115**, 186–191 (2006).
27. Dequiedt, S. *et al.* Biogeographical patterns of soil bacterial communities. *Environ. Microbiol. Rep.* **1**, 251–255 (2009).
28. Legendre, P. lmodel2: Model II Regression <http://CRAN.R-project.org/package=lmodel2> (2011).
29. Legendre, P. & Legendre, L. *Numerical Ecology* 2 edn (Elsevier Science BV, Amsterdam, 1998).
30. Bell, T. *et al.* Larger islands house more bacterial taxa. *Science* **308**, 1884 (2005).
31. Woodcock, S., Curtis, T.P., Head, I.M., Lunn, M. & Sloan, W.T. Taxa-area relationships for microbes: the unsampled and the unseen. *Ecol. Lett.* **9**, 805–812 (2006).
32. MacArthur, R.H. & Wilson, E.O. *The Theory of Island Biogeography* (Princeton University Press, Princeton, NJ, 1967).
33. Stearns, J.C. *et al.* Bacterial biogeography of the human digestive tract. *Sci. Rep.* **1**, 170–179 (2011).
34. Dequiedt, S. *et al.* Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. *Glob. Ecol. Biogeogr.* **20**, 641–652 (2011).
35. Hubbell, S. *A Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, NJ, 2001).
36. Zinger, L. *et al.* Contrasting diversity patterns of crenarchaeal, bacterial and fungal soil communities in an alpine landscape. *PLoS ONE* **6**, e19950 (2011).
37. Griffiths, R.I. *et al.* The bacterial biogeography of British soils. *Environ. Microbiol.* **13**, 1642–1654 (2011).
38. Fenchel, T. Biogeography for bacteria. *Science* **301**, 925–926 (2003).
39. Cho, J.-C. & Tiedje, J.M. Biogeography and degree of endemism of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66**, 5448–5456 (2000).
40. Gardi, C. *et al.* Soil biodiversity monitoring in Europe: ongoing activities and challenges. *Eur. J. Soil Sci.* **60**, 807–819 (2009).
41. Arrouays, D. *et al.* A new projection in France: a multi-institutional soil quality monitoring networkUne initiative nouvelle en France: la mise en place d'un réseau multi-institutionnel de mesure de la qualité des sols (RMQS). *C. R. Acad. Agric. Fr.* **88**, 93–105 (2002).
42. Quintana-Segui, P. *et al.* Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. *J. App. Meteorol.* **47**, 92–107 (2008).
43. Hill, M.O. & Smith, A.E.J. Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* **25**, 249–255 (1976).
44. Thioulouse, J. & Dray, S. Interactive multivariate data analysis in R with the ade4 and ade4TKGUI packages. *J. Stat. Softw.* **22**, 1–14 (2007).
45. Webster, R. & Oliver, M.A. *Geostatistics for Environmental Scientists* 2nd edn (Wiley, 2007).
46. Roberts, D.W. labdsv: Ordination and Multivariate Analysis For Ecology. <http://CRAN.R-project.org/package=labdsv> (2010).
47. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Development Core Team, Vienna, Austria, 2011).

## Acknowledgements

This study was supported by ADEME (Energy and Environment Management Agency) and by the French National Research Agency (ANR Biodiversity, ECOMIC-RMQS). RMQS soil-sampling and physico-chemical analyses were supported by a French Scientific Group of Interest on soils: the 'GIS Sol', involving the French Ministry for Ecology and Sustainable Development, the French Ministry of Agriculture, the French Agency for Energy and Environment (ADEME), the French Institute for Research and Development (IRD), the National Institute for Agronomic Research (INRA) and the National Institute of the Geographic and Forest Information (IGN). We thank all the soil surveyors and technical assistants involved in sampling the sites and the technical support from the French soil sample archive, which provided the studied collection of soils. We also thank D. Warwick and T.G. Orton for their assistance on the manuscript.

## Author contributions

L.R. managed the mathematical strategy and wrote the manuscript with N.C.P.B. N.C.P.B., S.D., J.T. and N.P.A.S. developed the strategy for mathematical data analysis and for the different ecological hypotheses. M.L. contributed to the molecular characterization of bacterial communities. P.A.M. participated to the development of the project and provided insightful comments and assistance in the redaction of the manuscript. F.E.R.M. developed and managed the MicroSol database required to organize and analyze the environmental and microbial metadata. C.J. coordinated the sampling campaigns on the French territory and managed the associated soil archive and data library. A.B., D.A. and P.L. permitted the scientific link between the RMQS sampling strategy and molecular microbial ecology by initiating scientific and technical collaborations.

## Additional information

**Competing financial interest:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Ranjard, L. *et al.* Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat. Commun.* **4**:1434 doi: 10.1038/ncomms2431 (2013).