Global drivers and patterns of microbial abundance in soil

Hector M. Serna-Chavez1,2†, Noah Fierer3 and Peter M. van Bodegom1*

1Department of Ecological Science, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands, 2Department of Life Sciences, University of Coimbra, P3004-517 Coimbra, Portugal, 3Department of Ecology and Evolutionary Biology and the Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA

ABSTRACT

Aim While soil microorganisms play key roles in Earth’s biogeochemical cycles, methodological constraints and sparse data have hampered our ability to describe and understand the global distribution of soil microbial biomass. Here, we present a comprehensive quantification of the environmental drivers of soil microbial biomass.

Location Global.

Methods We used a comprehensive global dataset of georeferenced soil microbial biomass estimates and high-resolution climatic and soil data.

Results We show that microbial biomass carbon (C_Mic) is primarily driven by moisture availability, with this single variable accounting for 34% of the global variance. For the microbial carbon-to-soil organic carbon ratio (C_Mic/C_Org), soil nitrogen content was an equally important driver as moisture. In contrast, temperature was not a significant predictor of microbial biomass patterns at a global scale, while temperature likely has an indirect effect on microbial biomass by influencing rates of evapotranspiration and decomposition. As our models explain an unprecedented 50% of the global variance of C_Mic and C_Mic/C_Org, we were able to leverage gridded environmental information to build the first spatially explicit global estimates of microbial biomass and quantified the global soil microbial carbon pool to equal 14.6 Pg C.

Main Conclusions Our unbiased models allowed us to build the first global spatially explicit predictions of microbial biomass. These patterns show that soil microbial biomass is not primarily driven by temperature, but instead, biomass is more heterogeneous through the effects of moisture availability and soil nutrients. Our global estimates provide important data for integration into large-scale carbon and nutrient models that may imply a major step forward in our ability to predict the global carbon balance, now and in a future climate.

Keywords Biogeochemical cycles, global carbon cycling, moisture limitation, nitrogen limitation, soil microbial abundance, soil microbial biomass, soil microbial carbon.

INTRODUCTION

Ecologists have long been intrigued by the distribution of life on Earth. Information on the distribution and abundance of aboveground plant biomass has been available for decades which improved our understanding of the processes structuring plant communities, driving species abundances and influencing ecosystem processes. In contrast, efforts to estimate global patterns in abundance and the size of the soil microbial biomass pool have lagged considerably. Given the key role soil microorganisms play in ecosystem processes, elucidating such patterns will represent a critical step as we seek to improve our ability to predict and understand microbial controls on biogeochemical cycles.
Local- and regional-scale studies have identified environmental drivers of microbial biomass carbon (C_{Mic}) and of the biologically active fraction of the soil organic carbon pool (C_{SoC}/C_{SoG}) (Insam, 1990; Wardle, 1992; Franzluhbbers et al., 2001; Bachar et al., 2008; Sinsabaugh et al., 2008). Fierer et al. (2009) presented results suggesting that the variability in C_{Mic} and soil information, we derive and present the first spatially consequent applying our multivariate models to gridded climate biomass and its relationship to climate and soil drivers. By sub-quantifying and explaining global patterns of soil microbial communities. Also, the interactions between above- and below-ground communities are increasingly understood; however, it is still unknown if the spatial patterns of soil microbial abundance mirror those of plant biomass. Therefore, we focus here on quantifying and explaining global patterns of soil microbial biomass and its relationship to climate and soil drivers. By subsequently applying our multivariate models to gridded climate and soil information, we derive and present the first spatially explicit global estimates of C_{Mic} in the soil profile and the topsoil C_{Mic}/C_{SoG} ratio. Our spatially explicit estimates of the soil microbial pool can substantially improve predictions of global biogeochemical and vegetation models and, thus, allowing better estimations of the carbon balance.

**MATERIALS AND METHODS**

**Soil C_{Mic} estimates**

We based our study on a dataset of soil microbial biomass estimates across all major biomes (Cleveland & Liptzin, 2007, as extended and modified by Fierer et al., 2009). To improve latitudinal representation, we complemented the dataset with estimates from arctic and tropical regions. The final dataset comprised 414 georeferenced estimates (in mg C_{Mic} kg^{-1} soil), most of which had been obtained from mineral soils and only a few from organic soils. All estimates were from the A horizons obtained under control (non-manipulated) conditions. No estimates from litter layers were included, as drivers of microbial biomass in litter layers are presumably different from those in soil. Also, because seasonal variation in microbial biomass was seldomly reported, we only present annual mean microbial biomass. The majority of estimates (92%) had been obtained using the chloroform fumigation-extraction (CFE) method (Vance et al., 1987) and a minor proportion (8%) from chloroform fumigation-incubation (CFI) (Jenkinson & Powlson, 1976). For both, microbial biomass in the depth-interval sampled was calculated as:

\[ C_{Mic} = E_C/k_{EC} \]  

where \( E_C \) represents the difference between fumigated and unfu- 
migated soil extractable carbon or CO₂ produced; and \( k_{EC} \) the extraction efficiency accounting for incomplete mineralization of biomass (CFI) or incomplete extraction of the killed biomass (CFE). While CFE and CFI are subject to biases in soils with high organic matter content and high acidity, respectively (Martens, 1995), no other method to quantify microbial biomass is bias-
free or as widely used (Martens, 1995; Joergensen et al., 2011). Moreover, we found no systematic deviations in microbial biomass estimates under potentially biasing conditions for CFE or CFI (see Appendix S1).

To facilitate comparisons across studies, we calculated \( E_C \) using the reported C_{SoG}, the applied extraction coefficient (k_{EC}) and equation 1. Afterwards, a k_{EC} of 0.4 was applied to stand-
ardize all estimates. In addition, considering soil bulk density and the depth interval sampled, we extrapolated microbial biomass content to the first metre of soil profile (C_{Mic,1} in g C \ m^{-2}). This extrapolation was made assuming that soil condi-
tions change continuously and similarly across all soils down to the first metre (Fierer et al., 2009) (Appendix S1). Because bulk density was usually not given for the individual studies, we applied the bulk density reported in the global database for the specific soil type (Batjes, 1995) if the estimate was from a mineral soil. Given that the bulk density of organic soils varies much more strongly, we used the value reported in the original publication (or from other publications describing the same site). In addition, because all depth intervals sampled for microbial biomass represented soil conditions in the upper 30 cm, we calculated the topsoil C_{Mic}/C_{SoG} considered an index of microbial activity. First, we normalized all estimates in the dataset to the amount of microbial biomass (g C_{Mic} m^{-2}) in the topsoil (Appendix S1). Then for mineral soils, we extracted topsoil SOC from the soil profile database (Batjes, 1995), whereas for organic soils, SOC was obtained from the original publication.

Global Ecology and Biogeography, ••, ••••, © 2013 John Wiley & Sons Ltd
Soil data

We used the ISRIC-WISE 0.5°-lat × 0.5°-long global database of soil profiles (Batjes, 1995) to extract soil variables, including bulk density (g cm⁻³), SOC (% mass), cation exchange capacity (CEC, meq 100 g⁻¹), C-N ratio (g C g⁻¹ N), pH (in H₂O solution), total nitrogen (% mass) and texture variables (% of sand and clay). These parameters reflect the physical-chemical characteristics of mineral soils (0–30 cm). Parameters for organic soils (0–30 cm) were extracted from the original publications or from related publications. Most commonly, studies reported SOC, total nitrogen, bulk density and pH. The missing parameters for organic soils were excluded from the analysis. These actions were a pragmatic solution to minimize biases because of missing organic soil data. Not considering microbial biomass estimates from organic layers may have led to underrepresentation or misrepresentation of particular global regions.

We used topsoil parameters to assess environmental drivers of soil profile CMic because they had higher correlation coefficients and accounted for more of the global variation in microbial biomass. We also ran analyses calculating soil parameters as (1) arithmetic mean of topsoil (0–30 cm) and subsoil (30–100 cm) data as reported in the ISRIC-WISE database, and (2) as weighted-means according to their relative importance to microbial biomass (87% topsoil, 13% subsoil). In both cases, neither correlation coefficients nor the explanatory power of multivariate models was higher than when using topsoil parameters alone.

Climate data

We used the climate research unit 0.5°-lat × 0.5°-long surface mean monthly climatology database of global land areas excluding Antarctica (New et al., 1999). Using the georeferences of the biomass sampling locations, we calculated mean annual precipitation (MAP, mm H₂O yr⁻¹); maximum monthly frequencies of wet and frost days (MaxWD and MaxFD, respectively), lowest annual minimum monthly temperature (MinAT, in °C), highest annual maximum monthly temperature (MaxAT, °C), mean annual temperature (MAT, °C), and annual temperature range (T_range = MaxAT – MinAT, in °C). MaxWD is defined as the highest monthly number of days where precipitation exceeded 1.0 or 0.1 mm. Similarly, MaxFD represents the highest monthly number of days with gross minimum temperature below 0°C. MinAT and MaxAT are defined as the lowest minimum and highest maximum monthly temperatures in a year, respectively. Additionally, we calculated mean monthly and annual evaporative demand (ETO, mm H₂O day⁻¹ and mm H₂O yr⁻¹), mean annual soil moisture deficit (MASMD, mm H₂O) and the annual ratio of MAP and ETO (MAP/ETO) for each biomass estimate. MASMD is the sum of differences between mean monthly ETO and mean monthly precipitation (Fierer & Jackson, 2006). ETO was calculated according to the Food and Agriculture Organization Penman–Monteith equation using climate data and assuming constant vegetation and roughness attributes, which makes the estimate independent of the standing vegetation (Allen et al., 1998).

Statistical analysis

Multivariate regression analyses were performed to quantify global environmental drivers and to model soil microbial biomass as a linear function of multiple explanatory variables. Estimates of soil profile CMic (g CMic m⁻²) and topsoil CMic/CORG ratio were used as response variables and analysed individually. Log₁₀ transformations were applied to both response variables to approach normal data distribution. Subsequently, all soil and climate variables were plotted against the transformed response variables to determine which to include in the regression analysis. In each case, Pearson’s correlation coefficients (r, α = 0.05) were calculated. We log₁₀-transformed some soil and climate variables to approach normal data distribution, enhance r, and improve linear relations between the explanatory and response variables. These transformations were performed separately for each analysis, as both response variables showed different patterns with explanatory variables. Upon the log₁₀ transformations, essentially linear relations were obtained.

To identify and select the most parsimonious set of environmental variables that explain CMic and CMic/CORG variance, we performed one-by-one backward stepwise regressions. To manage multicollinearity between environmental variables, we only included in our models explanatory variables of which the r among them varied between 0.7 and –0.7. Additionally, variance inflation factors (VIFs) were calculated for all regressors in multivariate models, and a limit of VIF ≤ 4 was set. Criteria to drop variables were: collinearity with other explanatory variables, significance of regression coefficients, single-term deletion F-test and the overall improvement of model’s fit after removal. The best multivariate models were selected according to the adjusted coefficient of determination (R_adj²), Akaike’s information criteria (AIC) and regression diagnostic plots (studentized residuals versus fitted values and distribution of studentized residuals). For the developed models, we assessed the relative variance accounted for by each explanatory variable using a hierarchical partitioning metric that accounts for the direct effect of each regressor and the adjusted effect for all other regressors (Img metric, Groemping, 2006).

As a control, we developed a null model to explain soil profile CMic and topsoil CMic/CORG using ‘biome’ as sole explanatory variable. All estimates in our dataset were classified as boreal forest, desert, temperate coniferous forest, temperate deciduous forest, temperate grassland, tropical forest or tundra. This simple biome model served to analyse interbiome differences and to compare further multivariate models with, that is, to assess whether or not we improved the proportion of explained variance by using continuous relationships with, environmental variables. The proportion of explained variance by interbiome differences alone was calculated as:

\[ \eta^2 = \left( \frac{SS_{\text{biome}}}{SS_{\text{biome}} + SS_{\text{residuals}}} \right) \times 100 \]
Furthermore, to cross-validate the developed multivariate models we compared our estimates to a dataset of substrate-induced respiration (SIR) estimates (μg C g⁻¹ h⁻¹) from 77 soil samples. This dataset comprised soil samples from an array of temperate soils in North America and a few tropical soils in Puerto Rico, as described in Fierer & Jackson (2006). SIR estimates were obtained for the top 30 cm of mineral soil using the methodology described in Fierer et al. (2003). For each location where SIR had been measured, we predicted soil profile C_Mic based on its georeferenced data to derive soil and climate variables using the 0.5°lat × 0.5°-long global soil profile (Batjes, 1995) and climate (New et al., 1999) databases, and our multivariate models. We then assessed the significance of correlations (Pearson’s r, α = 0.05) between our C_Mic estimates and the SIR-responsive biomass data. Finally, we mapped the global soil C_Mic and C_Mic/C_Org ratio applying our multivariate models to the gridted climate and the soil profile information for each 0.5° × 0.5° grid, following appropriate back-transformations. All statistical analyses were conducted in R v.2.13.0 (R Development Core Team, 2011), while the application of the developed regression models to derive the world maps was conducted in Fortran using Force 2.0 v.2.0.9p.

RESULTS
Quantification of global environmental drivers

Overall, moisture variables were the main individual drivers of soil profile C_Mic. Correlation with the MAP/ETo ratio (MAP/ETo) was the strongest (r = 0.46, P < 0.0001) followed by MAP (r = 0.45, P < 0.0001), both common measures of moisture availability and supply. Also, the interaction between moisture availability and MAT (MAP/ETo × MAT) exhibited a strong linear relationship with C_Mic (r = 0.57, P < 0.0001). This interaction indicated that with increasing MAT, the linear relationship between moisture availability (MAP/ETo) and microbial biomass becomes stronger. From our temperature-related variables only annual T_range had a strong correlation with C_Mic (r = −0.33, P < 0.0001). All other variables, including MAT, exhibited low correlation coefficients. Soil characteristics also showed an important influence; while the relation between SOC and C_Mic was not as strong as expected (r = 0.14, P = 0.004), total nitrogen and pH were quite strongly correlated to C_Mic (r = 0.25 and −0.26, respectively, P < 0.0001).

The analysis of topsoil C_Mic/C_Org data suggests that soil and temperature variables exert a stronger influence on this ratio than on C_Mic. Still, variables associated with soil moisture availability had the strongest correlation coefficients. However, it was MAP instead of MAP/ETo × MAT that exhibited the strongest correlation (r = 0.43 and 0.32, respectively, P < 0.0001). The influence of nutrients was exemplified by the strong correlation with total nitrogen content (r = −0.37, P < 0.0001) and, to a lesser extent, with the carbon-nitrogen (C-N) ratio (r = −0.25, P < 0.0001). MAT and the MinAT exhibited strong correlation with C_Mic/C_Org (r = 0.33, P < 0.0001, in both cases). Also, the MaxFD strongly constrained C_Mic/C_Org (r = −0.36, P < 0.0001). For both microbial biomass variables, all other environmental correlates are presented in Appendix S2.

Multivariate models to predict global patterns of microbial biomass

We selected the best multivariate linear models that explained soil profile C_Mic and topsoil C_Mic/C_Org (Table 1). Selected models had a distribution of studentized residuals close to normal, with no distinct patterns between residuals and the explanatory variables in the models or those variables not included in the model. Also, no true outliers were detected (Cook’s D < 0.08; Hat-values < 0.09; Bonferroni’s adjusted P > 0.05).

Soil profile C_Mic was best explained by a combination of MAP/ETo × MAT, the MaxWD, soil pH and total nitrogen, accounting for 39.0% of the total variance (Model 1, F_6,360 = 63.2, P < 0.0001, AIC_M = −875.6; Fig. 1a). A relative importance analysis of regressors supported the finding that moisture availability variables are the main drivers of C_Mic alone accounting for more than 80% of the explained variance (Fig. 2a). For topsoil C_Mic/C_Org the best model included MAP/ETo × MAT, MaxWD, soil total nitrogen, pH, C-N ratio and CEC. It explained 50.7% of the total variance in C_Mic/C_Org (Model 4, F_6,360 = 63.7, P < 0.0001; AIC_M = −802.9; Fig. 1b). Here, the relative importance analysis suggested a more balanced contribution between moisture availability and soil nutrient variables as main drivers, together accounting for more than half of the explained variance (Fig. 2b). Alternative multivariate models 2 and 3, representing the second best fits, are presented in Appendix S1.

For comparison, we also assessed interbiome differences for our dataset (see Methods). Biome level means of C_Mic were significantly different (ANOVA F_6,407 = 10.6, P < 0.0001) particularly because of tropical forests exhibiting higher microbial biomass than several other biomes (Fig. 3a). A similar result was obtained for the biome level means of C_Mic/C_Org (ANOVA F_6,390 = 11.2, P < 0.0001; Fig. 3b). These interbiome differences reflect inherent differences in environmental conditions among biomes and account for 13.5% (7–19%) and 14.7% [8–20%; ƞ² and 95% confidence interval (CI)] of the total C_Mic and C_Mic/C_Org variation, respectively. This level of variance is about three times less than the variance explained when directly coupling environmental drivers to biomass estimates. By directly estimating the impacts of environmental drivers on soil microbial biomass, we were able to account for the interbiome and intrabiome differences in microbial biomass.

We cross-validated our model estimates using microbial biomass estimated using a different methodology: SIR (μg C g⁻¹ h⁻¹; see Methods). Predicted C_Mic and SIR were better related to estimates from our best model (r = 0.39; P < 0.001) than to our alternative second-best model (r = 0.34; P < 0.05), although correlation coefficients were not statistically different from each other (Z_calculated = 0.35; P > 0.05). Given that this cross-validation compares local SIR estimates with grid-average microbial biomass, these correlation coefficients are reasonably
Table 1 Environmental drivers of soil profile CMic (Models 1 and 2) and the topsoil CMic/COrg ratio (Models 3 and 4)

<table>
<thead>
<tr>
<th>Drivers</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-value</td>
<td>β (± 1SEM)</td>
<td>b</td>
<td>P</td>
</tr>
<tr>
<td>MAP/ET o† × MAT</td>
<td>230.0</td>
<td>0.042 (± 0.004)</td>
<td>0.61</td>
<td>***</td>
</tr>
<tr>
<td>MaxWD</td>
<td>4.8</td>
<td>0.013 (± 0.005)</td>
<td>0.15</td>
<td>**</td>
</tr>
<tr>
<td>pH</td>
<td>14.1</td>
<td>0.090 (± 0.022)</td>
<td>0.21</td>
<td>***</td>
</tr>
<tr>
<td>Total N†</td>
<td>4.5</td>
<td>0.131 (± 0.062)</td>
<td>0.09</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>237.3</td>
<td>0.039 (± 0.004)</td>
<td>0.57</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>7.4</td>
<td>0.020 (± 0.005)</td>
<td>0.23</td>
<td>***</td>
</tr>
<tr>
<td>pH</td>
<td>11.4</td>
<td>0.071 (± 0.026)</td>
<td>0.17</td>
<td>**</td>
</tr>
<tr>
<td>C-N ratio†</td>
<td>20.7</td>
<td>-1.353 (± 0.432)</td>
<td>-0.14</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>0.008 (± 0.004)</td>
<td>0.10</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>131.4</td>
<td>0.652 (± 0.082)</td>
<td>0.46</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>24.9</td>
<td>-0.471 (± 0.129)</td>
<td>-0.18</td>
<td>***</td>
</tr>
<tr>
<td>MaxWD</td>
<td>0.5</td>
<td>0.023 (± 0.006)</td>
<td>0.24</td>
<td>***</td>
</tr>
<tr>
<td>Total N†</td>
<td>113.2</td>
<td>-0.817 (± 0.103)</td>
<td>-0.50</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>61.4</td>
<td>-2.943 (± 0.448)</td>
<td>-0.28</td>
<td>***</td>
</tr>
<tr>
<td>pH†</td>
<td>21.5</td>
<td>1.227 (± 0.343)</td>
<td>0.19</td>
<td>***</td>
</tr>
<tr>
<td>CEC†</td>
<td>4.6</td>
<td>0.376 (± 0.176)</td>
<td>0.12</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>183.5</td>
<td>0.041 (± 0.004)</td>
<td>0.55</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td>0.022 (± 0.005)</td>
<td>0.23</td>
<td>***</td>
</tr>
<tr>
<td>Total N†</td>
<td>225.0</td>
<td>-0.961 (± 0.095)</td>
<td>-0.59</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>43.1</td>
<td>-2.067 (± 0.441)</td>
<td>-0.19</td>
<td>***</td>
</tr>
<tr>
<td>pH†</td>
<td>24.1</td>
<td>1.268 (± 0.334)</td>
<td>0.20</td>
<td>***</td>
</tr>
<tr>
<td>CEC†</td>
<td>5.2</td>
<td>0.396 (± 0.173)</td>
<td>0.12</td>
<td>*</td>
</tr>
</tbody>
</table>

*P < 0.05, **P < 0.001, ***P < 0.001.
†Variable is log10-transformed.
β, regression coefficients; b, standardized coefficients; CEC, soil cation exchange capacity; CMic, microbial biomass carbon; CMic/COrg, microbial carbon-to-soil organic carbon ratio; C-N ratio, soil carbon-nitrogen ratio; ETo, annual evaporative demand; MAP, mean annual precipitation; MAT, mean annual temperature; MaxWD, maximum monthly frequency of wet days; pH, soil pH; SEM, standard error of the mean; Total N, soil total nitrogen content; VIF, variance inflation factors.

Figure 1 Observed versus predicted values. (a) Soil profile microbial biomass carbon (CMic) estimates from Model 1 (as provided in Table 1). Regression line: slope of 1.000 ± 0.062; β ± 1 standard error of the mean (SEM)]; R² of 39.8%, P < 0.0001 (n = 391). (b) Topsoil CMic/soil organic carbon (COrg) ratio estimated using Model 4 (as provided in Table 1). Regression line: slope of 1.000 ± 0.051; β ± 1SEM); adjusted coefficient of determination (Radj²) of 50.7%, P < 0.0001 (n = 367). For both panels, shadowed area represents the 95% confidence interval (CI), and all axes are on a log10 scale. Legend and biome classification as follows: boreal forest (BF), desert (D), tundra (T), temperate coniferous forest (+, TCF), temperate deciduous forest (x, TDF), tropical forest (∇, TF), and temperate grassland (∇, TG).
More importantly, the residual variance of the cross-validation did not show deviations from null patterns, indicating unbiased estimates.

Global estimates of microbial biomass

The spatially explicit global patterns of microbial biomass derived from our multivariate models are presented in Fig. 4. Hotspots of soil profile CMic were in tropical regions; however, comparable abundances were evident in some subtropical, temperate and boreal regions. The lowest CMic was estimated for arid and semi-arid regions (Fig. 4a). Our multivariate models led to an estimated global soil profile microbial biomass pool of 14.6 Pg CMic (95% CI). For the topsoil CMic/soil organic carbon (COrg) ratio, (Radj^2 = 50.7%). In both cases, metrics have been normalized to sum 100%. Confidence intervals (95%) were obtained after 1000 bootstrap runs. Equal lower-case letter indicate that confidence intervals for the difference between the contributions of each regressor includes zero (P < 0.05). Abbreviations as explained in Table 1.

DISCUSSION

Global environmental drivers of microbial abundance

By analysing the continuous relations between environmental drivers and microbial biomass, we were able to explore the variations between and within biomes without any a priori classification and reduced the residual variance in biomass estimates. In our analysis, moisture-related variables exhibited the strongest linear relationships with CMic and CMic/COrg. This may be no surprise, as moisture availability can strongly affect microbial activity in both wet and dry climates (Bachar et al., 2010; Blankinship et al., 2011) while also influencing nutrient availability. Although high soil moisture is known to restrict microbial activity (Drenovsky et al., 2010; Wu et al., 2011), we found no evidence for such limitations, likely because anaerobiosis is sufficiently infrequent to limit microbial biomass in surface soils. Furthermore, because both CMic and CMic/COrg were positively related to annual moisture availability, this suggests that microbial biomass increases more with precipitation than SOC. By including the inherent influence of temperature on annual moisture availability (MAP/ETo X MAT interaction) we were able to better explain the variability in microbial biomass.
Because MAT alone showed little correlation with CMic, while annual ETo, MAP/ETo and the MASMD were all strongly correlated (Appendix S2), this suggests that temperature predominantly affects CMic indirectly by influencing evapotranspiration rates. Conversely, both temperature variables and moisture were important drivers of CMic/COrg, in contrast with effects found for CMic. This suggests that the influence of temperature on CMic/COrg is mainly through its effects on SOC, which is known to be strongly dependent on temperature (Parton et al., 2007). Furthermore, any variable that expressed the seasonality of precipitation (e.g. coefficient of variation or precipitation in the wettest quarter), as well as of temperature (e.g. temperature of the coldest quarter), was less able to account for the variance in either CMic or CMic/COrg than mean annual variables (Appendix S1).

Soil parameters were also important drivers of CMic and especially of CMic/COrg suggesting that soil conditions have a stronger influence on the amount of organic carbon immobilized in microbial biomass than on biomass itself. The positive relationships we found between CMic, SOC and total nitrogen content are in line with previous evidence (Wardle, 1992; Cleveland & Liptzin, 2007; Fierer et al., 2009). Multivariate models that included total nitrogen instead of SOC explained a higher variance of CMic and were a better fit despite total nitrogen being highly correlated with SOC ($R^2 = 95\%$; $F_{1,392} = 7281$; $P < 0.0001$). Thus, while the relation between CMic and total nitrogen also reflects an influence of SOC, our results suggest that effects of nitrogen limitation are more consistent. However, these effects seem to decrease at high substrate concentrations because low microbial biomass was found in tundra soils at high total nitrogen (and coinciding high SOC; Appendix S2). These areas are characterized by climatic stress, such as prolonged low temperatures and long periods of moisture shortage because of frost, which can restrict decomposition leading to accumulation of organic matter (Wardle, 1992; Cleveland & Liptzin, 2007; Fierer et al., 2009). Also, this accumulated carbon may be more recalcitrant (lower substrate quality; Franzluebbers et al., 2001; Fierer et al., 2009). As depicted in our maps, low CMic/COrg ratios are evident for these regions under climatic stress, while the estimated CMic is similar to that found in tropical regions.

Unlike for CMic, the effect of soil nitrogen on CMic/COrg was negative. Given that CMic/COrg is generally considered an index of microbial activity, these results suggest that in addition to limitations by soil nitrogen on microbial biomass, nitrogen may also hamper microbial activity. A biome-deviation analysis revealed that for all biomes, the relationship was indeed negative except for desert soils (see Appendix S1). The latter is expected because in arid regions, soil nitrogen is considered a limiting resource (Gallardo & Schlesinger, 1992). Thus, increased nitrogen availability could allow for an increase in CMic/COrg if other environmental conditions are favourable. A negative relationship does not seem to be due to impaired decomposition by anoxic conditions with accumulated soil organic matter (as occurs in e.g. humid tropical forest, Pett-Ridge et al., 2006), given that this cannot explain the negative impacts in other biomes. However, the negative relationship is consistent with nitrogen fertilization.
This may be explained by a shift in the soil microbial community to one with lower standing biomass but higher turnover rates given higher nutrient availabilities (Allison et al., 2008; Ramírez et al., 2010) or a reduction in the allocation of belowground labile carbon with increasing soil nitrogen contents (Dijkstra et al., 2005; Allison et al., 2008; Eisenhauer et al., 2012), which may subsequently increase the carbon demand of microbial biomass and induce shifts in community structure (Allison et al., 2008). Given that aforementioned processes are intimately linked, they may represent complementary responses (with potentially different importance to subsets of our data) and may together help explain the overall negative global relationship between total nitrogen and $\text{CMic}/\text{COrg}$.

Soil resource quality (C-N ratio) also exhibited an important influence on both microbial variables. High C-N ratios tend to indicate low organic matter decomposability, whereas higher substrate quality enhances microbial activity and growth (Parton et al., 2007; Dequiedt et al., 2011). Also, environmental stress can limit decomposition causing an accumulation of soil carbon and a higher C-N ratio because of accumulation of undecomposed plant litter. This might restrain microbial biomass abundance by reducing substrate quality. These mechanisms account for the negative relationship between our response variables and soil C-N ratio. Furthermore, according to our models, $\text{CMic}/\text{COrg}$ and $\text{CMic}$ increased with pH. Acidic conditions are known to restrict organic matter decomposition and microbial (enzymatic) activity. As pH increases, microbial activity follows, allowing the immobilization of a bigger proportion of organic carbon in microbial biomass (Wardle, 1992; Sinsabaugh et al., 2008). In accordance with previous studies (Anderson & Domsch, 1993; Baath & Anderson, 2003), this can translate into an overall positive effect of pH on microbial biomass. Soil texture variables also exhibited important influence on $\text{CMic}/\text{COrg}$ and on $\text{CMic}$ (Appendix S2). It is likely that the relation between CEC and $\text{CMic}/\text{COrg}$ reflects an enhanced retention of organic matter and reduced quality (Wardle, 1992; Six et al., 2002).

The patterns mentioned earlier might be the result of complex interacting effects in the soil environment and with climate (e.g. the climate, soil organic matter and pH relationship; Insam, 1990; Jobbagy & Jackson, 2000; Franzluebbers et al., 2001). Nonetheless, the drivers described here are likely the main actors because of the imposed restraints on collinearity between environmental variables in our models (see Methods).

**Spatial patterns and estimates of microbial biomass**

Our unbiased multivariate models allowed us to build global spatially explicit predictions of microbial biomass in soils. These patterns show that soil microbial biomass is not primarily driven by temperature (as reflected in latitude), but instead biomass is more heterogeneous through the effects of moisture availability and soil nutrients (Fig. 4). Similarly, precipitation and soil nutrients have been shown to influence patterns of plant biomass (Moles et al., 2009; Vicca et al., 2012). To assess whether global patterns in plant and soil microbial abundance coincide, we tested the relation between net primary production of potential vegetation (NPP in g m$^{-2}$ year$^{-1}$, from the Lund-Potsdam-Jena-Dynamic Global Vegetation Model, Sitch et al., 2003) and soil $\text{CMic}$. Their correlation coefficient was comparable ($r = 0.5$, $P < 0.0001$) to those of annual moisture variables (Appendix S2). However, a large portion of the variance in NPP is already accounted for by annual moisture supply and availability variables (e.g. MAP and MAP/ETo $\times$ MAT, in bivariate regressions: $R^2 = 0.52$ and 0.45, respectively; $P < 0.0001$). Indeed, including NPP in our models did not increase the amount of explained variance or their fit. Instead, replacing the main driver (i.e. MAP/ETo $\times$ MAT for the $\text{CMic}$ model) with NPP led to a decrease in explained variance from $R^2 = 0.39$ ($P < 0.0001$; AIC$ = -876$), to $R^2 = 0.28$ ($P < 0.0001$; AIC$ = -816$). Compared with soil microbial biomass, plant biomass seems to be more constrained by temperature (Law et al., 2002), as MAT was more strongly correlated to NPP ($r = 0.39$; $P < 0.0001$) than to $\text{CMic}$ ($r = 0.07$; $P > 0.05$). Together, this seems to reveal that (1) the extent and direction of these relationships with environmental drivers is different for plant NPP and soil microbial biomass, and because our predictions accurately portrayed average observations (see 1:1 patterns in Fig. 1). And, in spite of this discrepancy in scale, our validation showed similar patterns to the observed SIR as reported earlier for local datasets (Anderson & Joergensen, 1997; Fierer et al., 2003, 2009). Considering uncertainties in our models, the correlation between our biomass estimates and SIR data are in line with previous studies (thus further supporting the contention that our estimates and spatial patterns are unbiased). Second, to standardize measurements, we extrapolated microbial biomass estimates from the depth-interval sampled to the first meter. This might have biased estimates for specific sites, e.g. where data were obtained from organic layers on top of mineral soils. However, even though depth distributions will differ among sites, and biomes, it presents a shape similar to the global average SOC depth distribution (Jobbagy & Jackson, 2000) and likely only contributed to residual variance in our models. Moreover, including soil parameters of subsoil conditions did not improve our models (see Methods). Future work elucidating microbial biomass depth distribution in different soil environments will help making suitable extrapolations and, thus, better comparisons. A better characterization of soil parameters along the profile might help improving the resolution of our multivariate models.
(2) that annual climatic variables are better able than NPP to account for the global variance in soil microbial abundance.

Our comprehensive analysis provides a global estimate of 14.6 Pg C_Mic in the soil profile. This estimate falls within the range of previous estimates (Wardle, 1992; Whitman et al., 1998; Xu et al., 2013), which had been hampered by low representation in particular biomes and lack of spatially explicit analysis of drivers. In addition, according to the well-conserved C : N : P ratio of microbial biomass (Cleveland & Liptzin, 2007), our findings suggest that the global soil microbial nitrogen and phosphorous pools are around 2.0 and 0.6 Pg, respectively. Our study also estimate the global topsoil C_Mic/C_Org ratio at 1.2% (0–2%; 75% CI), which is in line with previous studies that estimated it for topsoil (cumulatively; 0.3–5%; Insam, 1990; Wardle, 1992; Zak et al., 1994; Xu et al., 2013) and for the soil profile (0.6–1.1%; Fierer et al., 2009), respectively.

Microbial biomass estimates used in our study are mostly representative for the active season, which matters most for ecosystem fluxes and services. Thereto, our estimates are well suited to understand and to further improve quantitative estimates of ecosystem fluxes and services, now and in the face of global environmental change. In fact, our quantitative estimates may serve many applications as soil microorganisms mediate the rates of carbon and nutrient cycling in terrestrial ecosystems (Wardle, 1992; Falkowski et al., 2008). Consequently, microbial biomass serves as a critical control on the feedbacks between global change, organic matter decomposition and ecosystem productivity (Manzoni & Porporato, 2009; Todd-Brown et al., 2012). Current global biogeochemical and vegetation models have only a coarse, if any, representation of soil microbial properties and most commonly frame decomposition of organic matter as first-order decay, that is, only as a function of environment (temperature and moisture) and not of microbial biomass (Parton et al., 1993; Sitch et al., 2003), hampering the predictive ability of these models (Manzoni & Porporato, 2009; Todd-Brown et al., 2012). The problems with scaling up information on microbial abundance and activity from sparse local data (Todd-Brown et al., 2012) have made it difficult to integrate soil microbial biomass into global scale decomposition models. However, our global estimates of soil microbial biomass provide important data for integration into large-scale carbon and nutrient models, which may imply a major step forward in our ability to predict the global carbon balance, now and in a future climate. Thereto, our global maps of C_Mic and topsoil C_Mic/C_Org are available in Appendix S3.

In summary, our study shows how using average environmental variables allows us to explain a large portion of the global variance in soil microbial biomass. By identifying generic environmental constraints on soil microbial abundance, we show the stronger constraints exerted by soil variables and temperature on C_Mic/C_Org than on C_Mic, suggesting that topsoil microbial activity is further constrained by local soil conditions and low temperatures. Also, by deriving a spatially explicit global map of soil profile C_Mic and topsoil C_Mic/C_Org distribution, we provide an improved spatially explicit estimate of the global biological soil carbon pool. Our estimates, in combination with new understanding of substrate and decomposer stoichiometry (Cleveland & Liptzin, 2007; Manzoni et al., 2008; Sinsabaugh et al., 2008), will together improve our ability to predict the controls and rates of organic matter decomposition, nutrient mineralization and soil-atmosphere carbon exchange.

ACKNOWLEDGEMENTS

The author H.S.C. would like to thank the European Commission’s Education, Audiovisual and Culture Executive Agency (EACEA) for its support throughout the research period with an Erasmus Mundus Scholarship (2009–2011). Also, the authors would like to thank E. van Loon for comments on an earlier version of the manuscript.

REFERENCES

Anderson, T.H. & Domsch, K.H. (1993) The metabolic quotient for CO2 (QCO2) as a specific activity parameter to assess the effects of environmental conditions, such as pH, on the microbial biomass of forest soils. Soil Biology and Biochemistry, 25, 393–395.


R Development Core Team (2011) *R: a language and environment for statistical computing*. R Development Core Team.


**SUPPORTING INFORMATION**

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

**Figure S1** Secondary multivariate models that predict patterns of C_Mic and C_Mic/C_Org (A) Observed and predicted soil profile C_Mic estimates from Model-2: slope = 1.0 (± 5.9 x 10^{-2}, \( \beta \) ± 1SEM), intercept: -1.6 x 10^{-8} (± 1.2 x 10^{-7}), \( R^2 \) = 43.9%, \( F_{1,362} = 285; P < 0.0001 \). (B) Observed and predicted topsoil C_Mic/C_Org ratios from Model-3: slope = 1.0(± 5.2 x 10^{-2}; \( \beta \) ± 1SEM), intercept: -1.8 x 10^{-7}, \( R^2 \) = 49.9%, \( F_{1,362} = 363; P < 0.0001 \). Shadowed area is the 95% C.I and all axes are in log10 scale. Panels (C) and (D) are the relative importance of regressors included in Model 2 and 3, respectively. In both, panels equal lower-case letter indicate that confidence intervals for the difference between the contributions of each regressor includes zero (\( P < 0.05 \)). Also, metrics are normalized to sum 100% of explained variance and 95% C.I. were obtained after 1000 bootstrap runs. Legend and biome classification as follows: boreal forest (\( \square \), BF), desert (\( \bigcirc \), D), tundra (\( \triangle \), T), temperate coniferous forest (+, TCF), temperate deciduous forest (\( \times \), TDF), tropical forest (\( \bigodot \), TF), and temperate grassland (\( \nabla \), TG). MAP: mean annual precipitation; ETo: annual evaporative demand; MAT: mean annual temperature; MaxWD: maximum monthly frequencies of wet days; Total N: soil total nitrogen content; CEC: soil cation exchange capacity; pH: soil pH; C-N ratio: soil Carbon-Nitrogen ratio.

**Table S1** Environmental drivers of soil profile microbial biomass carbon (\( g_{C_{Mic}} m^{-2} \)). Bivariate regressions of annual and seasonal climatic variables versus microbial biomass.

**Appendix S1** Supplementary information on materials, methods, and results.

**Appendix S2** Scatterplots and Person’s correlation factors of both soil microbial biomass carbon (C_Mic) and soil microbial biomass carbon-soil organic carbon (C_Mic/C_Org), and environmental variables.

**Appendix S3** ASCII maps of soil profile microbial biomass carbon (C_Mic), and for the topsoil soil microbial biomass carbon-soil organic carbon (C_Mic/C_Org).

**BIOSKETCHES**

**Hector Serna-Chavez**’ main research interests focus on the underlying role of functional diversity on ecosystem services, including the spatiotemporal characteristics of delivery and its resilience under global environmental change.

**Noah Fierer**’s research focuses on the ecology and biogeography of microorganisms in a range of systems including the atmosphere, soil, and the human body, exploring linkages between microbial diversity and the functioning of microbial communities.

**Peter van Bodegom** works at the interface between community ecology, macroecology and earth system modelling. With his group, he aims to quantify globally applicable functional relationships between vegetation, soil microorganisms, and their environment by targeted experiments, meta-analyses and process-based modelling.