Environmental drivers of soil microbial community distribution at the Koiliaris Critical Zone Observatory

Myrto Tsiknia[a], Nikolaos V. Paranychianakis[a], Emanouil A. Varouchakis[a], Daniel Moraetis[a,b], Nikolaos P. Nikolaidis[a]

[a]Dept of Environmental Engineering, Technical University of Crete, Polytechnioupolis 73100, Chania Greece

[b]Sultan Qaboos University, College of Science-Earth Science Department P.O.Box 36, P.C. 123, Muscat

Keywords: soil microbial community, qPCR, critical zone, spatial distribution

Corresponding author: N.V Paranychianakis

[a]School of Environmental Engineering, Technical University of Crete, Polytechnioupolis 73100, Chania Greece

Tel.: +30 28210 37823; Fax: +30 28210 37846

Email: niko.paranychianakis@enveng.tuc.gr

Running title: Soil Microbes distribution at a watershed scale

Abstract

Data on soil microbial community distribution at large scales are limited despite the important information that could be drawn with regard to their function and the influence of environmental factors on nutrient cycling and ecosystem services. This study investigates the distribution of Archaea, Bacteria and Fungi as well as the dominant bacterial phyla (α- and β-Proteobacteria, Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes) across the Koiliaris watershed by qPCR and associate them with environmental variables. Predictive maps of microorganisms distribution at watershed scale were generated by co-kriging, using the most significant predictors. Our findings showed that 31–79% of the spatial variation in microbial taxa abundance could be explained by the parameters measured, with total organic carbon and pH being identified as the most important. Moreover, strong correlations were set between microbial groups and their inclusion on variance explanation models improved their prediction power. The spatial auto-correlation of microbial groups ranged from 309 to 2.226 m and geographic distance, by itself, could explain a high proportion of their variation. Our findings shed light on the factors shaping microbial communities at a high taxonomic level and provide evidence for ecological coherence and syntrophic interactions at the watershed scale.

Introduction

Microorganisms regulate the biogeochemical cycles of nutrients in terrestrial ecosystems and hence, the services provided. In turn, the prevailing environmental conditions, including both biotic and abiotic factors, exert an apparent control on microbial community structure and activity. An increasing body of literature indicates

that climate, soil properties, vegetation, and land use and management as important determinants of the abundance, structure, and activity of soil microbial community (King *et al.*, 2010; Nielsen *et al.*, 2010; Rousk *et al.*, 2010; Wessén *et al.*, 2010; Zinger *et al.*, 2011). The relationships between microbial community structure and function in response to environmental parameters and management practices have been poorly understood (Fierer *et al.*, 2007; Strickland *et al.*, 2009; Fierer *et al.*, 2013). Spatial data on environmental variables is envisaged to improve our understanding on evolutionary factors shaping microbial communities and mediating their function. In addition, there is an ongoing discussion whether the inclusion of data microbial community structure will improve the simulations of the (global) biogeochemical models (Allison *et al.*, 2010; Wieder *et al.*, 2013).

The study of soil microbial biogeography is an emerging research field and lacks behind biogeochemical data and/or physical properties. Spatial studies in soil microbial community structure have been carried out at various scales, ranging from soil pore (Ruamps *et al.*, 2011), to individual fields (Philippot *et al.*, 2009), regional (Bru *et al.*, 2011), country (Griffiths *et al.*, 2011), continental level (Lauber *et al.*, 2009; Fierer *et al.*, 2013), or global level (Fierer *et al.*, 2009; Nemergut *et al.*, 2011; Serna-Chavez *et al.*, 2013). The sampling density, the soil properties assessed (physical, chemical, biological), and the method and the depth of microbial community characterization diverge greatly among studies.

The employment of microbial taxa at high taxonomic levels to gain information on their ecological niches and to assign them functions has been questioned (Green *et al.*, 2008; Philippot *et al.*, 2010; Nemergut *et al.*, 2013). The enormous diversity of soil microbes encompassed (e.g. Proteobacteria), and hence the functional traits carried by them (Baldrian *et al.*, 2012), has been identified as the major constraint. There is also a

number of studies indicating that pertinent information at the phylum level could provide important information for the function and the ecological niches of soil microorganisms. For instance, bacterial and archaeal phyla responded variably to changes in soil management practices (Wessén *et al.*, 2010). Fierer *et al.* (2007) performing a sampling campaign in many field sites accompanied by a meta-analysis of published data classified some bacterial phyla as copiotrophs and oligotrophs, allowing us to make predictions about their aggregated ecological attributes. Probably most importantly, data on soil microbial structure even at a high taxonomic level may provide critical information on soil biogeochemical cycles and their modeling (Averill *et al.*, 2014). Information at the domain level, addressed by the fungi:bacteria ratio, improved model simulations in terms of C and N mineralization that mainly resulted from differences in bacterial and fungal physiology (Waring *et al.*, 2013).

In this study, we characterize the distribution of soil microorganisms at the domain and phylum level and provide insights on the environmental variables that drive their spatial variability at the scale of a watershed. The Koiliaris river watershed is a representative Mediterranean watershed that has been recently characterized as Critical Zone Observatory (CZO) (Banwart *et al.*, 2011). Moreover, the dense availability of data relevant to soil genesis and evolution, soil physico-chemical properties, land uses, agricultural practices, climatic variability and hydrology, constitute Koiliaris CZO an interesting case to interpret the effects of these parameters on microbial taxa abundance as well as to elucidate the drivers shaping soil microbial communities. The abundance of Archaea, Bacteria, Fungi and of major bacterial phyla was quantified in soil samples, collected across the Koiliaris CZO which extends to an area of approximately 130 km^2. The Koiliaris CZO is characterized by steep gradients in climatic conditions, soil pedology and geomorphology and variable land uses. Variance partitioning was applied

to explain the relative contributions of climate, land use, spatial distance, and eleven physical, chemical and biochemical soil properties to microorganisms distribution. Then, geostatistical modelling (co-kriging) was employed to investigate the spatial correlations of microbial groups and to generate distribution maps at the watershed scale. In a following step, we included microbial groups in co-kriging to check the hypothesis if there are any microbial group(s) at the domain or phylum level that could improve the predictions obtained. The findings of the present study improve our understanding on the environmental factors regulating the abundance and distribution of microorganisms at a watershed scale as well as their interrelationships.

Materials and methods

Study area description

Koiliaris CZO is located 25 km east of Chania city, Crete, Greece (005-12-489E, 039-22-112N). The watershed consists of soils depleted in soil organic carbon and severely degraded due to intense agricultural practices and over-grazing. The Koiliaris CZO occupies an area of approximately 130 km^2 and is characterized by an intense topography extending from sea level to 2,100 m. More details on Koiliaris CZO with regard to prevailing climatic conditions, pedology, soil evolution and land use can be found in Moraetis *et al.* (2011) and Moraetis *et al.* (2014).

Soil sampling

Composite samples (three soil cores from each sampling point) were taken from 0-15 cm soil depth during the period May 15 to June 3, 2012 (Fig. 1). Sampling points were carefully selected following field campaigns in order to effectively capture the great variability observed at the Koiliaris CZO with regard to climate, soil properties and

115 land uses. Samples were passed through a 2-mm mesh immediately after sampling at
116 the field and maintained on ice packs at 4 ^{o}C until they were transferred to the
117 laboratory. There, each sample was split into subsamples for chemical, biochemical and
118 biological assays. The later samples were stored at -80°C until the genomic DNA
119 extraction and biochemical assays were carried out. The coordinates of each sampling
120 point were recorded with a global positioning system device.

121 Soil physical, chemical, biochemical analyses

122 Soils moisture was determined by drying subsamples to constant weight at 65 ^{o}C.
123 Electrical conductivity (EC) and pH were measured in H_2O at a soil:solution ratio of
124 1:2.5. NH_4^+-N and NO_3^--N were extracted with 2 M KCl by shaking the samples for 30
125 min, and were measured colorimetrically in a Perkin-Elmer Lambda 25
126 spectrophotometer with the Nessler reagent and the Cd reduction method, respectively.
127 Total organic carbon (TOC) and total nitrogen (TN) were measured in a Analytik Jena
128 Multi N/C® 2100S analyzer. Particle size analysis was carried out by the Bouyoucos
129 hydrometer method (Bouyoucos, 1962).

130 Net N mineralization rate (NMR) and potential nitrification rate (PNR) assays were
131 employed to follow the mineralization of organic-N and its subsequent oxidation to
132 NO_3^--N. Both assays were assessed in triplicates immediately after sampling. PNR
133 assays were performed according to the method developed by Smolders *et al.* (2001)
134 with modifications (Tsiknia *et al.*, 2014). NMR was estimated with the laboratory
135 aerobic incubation method (Hart *et al.*, 1994).

136 The potential activity of urease (EC 3.5.1.5), phenol oxidase (EC 1.10.3.2), and
137 peroxidase (EC 1.11.1.7) were assessed according to the protocols of Kandeler &
138 Gerber (1988), Li *et al.* (2010), and Sinsabaugh *et al.* (2005), respectively, since they

were believed to have an important role on C and N cycling in Koiliaris CZO. More specifically, phenol oxidase activity is mainly attributed to Fungi although some groups of bacteria have also been reported to be involved (Theuerl & Buscot, 2010). Hence, this assay was selected as a proxy of fungal abundance and activity. Furthermore, the Koiliaris CZO is dominated by olive trees, the litter of which is characterized by high phenol content, and thus, we assumed that phenol oxidase may have an important role in C cycling in the study area. Peroxidases are extracellular enzymes with an important role on soil C cycling through the depolymerization of recalcitrant macro-molecules (Sinsabaugh, 2010). Finally, urease catalyzes the hydrolysis of urea in agricultural fields and overall urease has an important role in C and N cycling in terrestrial ecosystems (Kandeler *et al.*, 1999). Information on the environmental and biogeochemical parameters measured from all sampling points across the Koiliaris CZO is summarized in Table S1.

DNA extraction and quantitative PCR (qPCR) assays

Microbial genomic DNA was extracted in triplicates from 0.25 g of soil, previously frozen and homogenized with mortar, using the PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Inc. Carlsbad, CA, USA) according to manufacturer's instructions. The three DNA extractions per soil sample were pooled before further analysis. DNA quality from each sample was checked in agarose gel (1%) and quantified in a NanoPhotometer® Pearl (Implen) and stored at -80 ^{o}C. Amplification conditions and primer pairs used in this study to quantify Fungi, Archaea, Bacteria, α- and β-Proteobacteria, Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes are summarized in Table S2. Quantification of gene copy numbers was carried out with the StepOnePlus™ Real-Time PCR System (Applied Biosystems) in 20 μl reactions using the KAPA SYBR Fast Master Mix (2x) qPCR Kit (KAPA Biosystems) and 0.8 – 3.5

ng of DNA. All reactions were completed with a melting curve starting at 60 ^{o}C, with an increase of 0.5 ^{o}C, up to 95 ^{o}C to verify amplicon specificity. Standard curves were obtained using serial dilutions, 10^{3}-10^{7}, of linearized plasmids (pGEM-T, Promega) containing cloned genes from each domain/phylum. Controls without template resulted in undetectable products in all the samples, while inhibitory effects were not detected at the dilution used (1/10). The amplification efficiencies ranged from 80% to 92% and the R^2 values of the standard curves ranged from 0.993 to 0.999.

Statistical analysis

All variables, except pH, were prior transformed according to their skewness and kurtosis characteristics to meet the assumption of normality. The transformation applied in each of the variables is shown in Table S3. Pearson correlation was employed to determine the significance of the relationships between microbial taxa and biogeochemical parameters. The Principal Coordinate of a Neighbor Matrix (PCNM) approach (Borcard & Legendre, 2002) was applied to geographic coordinates of the sampling points to construct spatial vectors in accordance to earlier studies (Bru *et al.*, 2011). Twenty-seven spatial variables were constructed representing all spatial scales present in the study area. The most significant explanatory and PCNM variables were selected by stepwise multiple regression analysis to predict the abundance of microbial domains or phyla. In order to determine the unique variance explained by each predictor, the final R-squared model was partialled out, based on the theory of squared semi-partial correlation (Legendre & Legendre, 1998) by adjusting R^2 values (% of explained variation) to obtain unbiased estimates (Peres-Neto *et al.*, 2006). Bonferroni –correction was applied to p values to maintain the family-wise error level in multiple testing. Analysis of Variance (ANOVA) was carried out between microbial groups and land use, separated in two broad categories agricultural land and natural ecosystems,

and also with elevation, separated in three groups (0 to 200 m; 201 to 400 m; and > 400 m). Homoscedasticity was verified using the Levene's test. All statistical analyses were performed with the R statistical platform using the vegan, MASS and yhat packages (Venables & Ripley, 2002; Nimon *et al.*, 2013; Oksanen *et al.*, 2013; R Development Core Team, 2013). The correlogram of the Pearson's correlation coefficients constructed with the corrgram package (Wright, 2013).

Geostatistical interpolation

The accurate mapping of microorganisms' abundance in soil is important for effective ecosystem management and monitoring decisions. Estimates at unsampled locations can be obtained by applying stochastic and deterministic interpolation methods to the available data (Deutsch & Journel, 1992; Goovaerts, 1997; Varouchakis & Hristopulos, 2013a). Stochastic methods such as kriging are commonly adopted since they allow the estimation of interpolation uncertainties (Deutsch & Journel, 1992). Optimal kriging results are obtained when the probability distribution of the data is jointly normal and stationary in space. Kriging estimates are linear combinations of the data with weights that follow from the no-bias constraint and the minimization of the mean square error. The weights are determined from a model semi-variogram, which is commonly obtained by fitting the empirical semi-variogram of the measurements to theoretical models. The semi-variogram measures the spatial correlation as a function of the distance between data points (Goovaerts, 1997). The theoretical semi-variogram model fitting is usually expressed by three parameters, the nugget that refers to the nonzero intercept due to measurement error or variation within the distance sampling interval, the sill that represents the variance of the correlated measurements and the range that defines the distance, extending from any given location, where measurements are spatially correlated (Goovaerts, 1997). Ordinary kriging (OK) is the most common

methodology that bases its estimates at unsampled locations only on the sampled primary variable (Kitanidis, 1997). Co-kriging (CoK) involves auxiliary variables that are significantly correlated with the primary variable leading to predictions with improved accuracy. CoK is a weighted average of measured values of the primary variable and of the cross-correlated auxiliary variables. The spatial correlation between two or more variables at the same location is expressed by the cross-semi-variogram (Kitanidis, 1997).

Semi-variogram modeling was initially performed in Matlab platform with codes developed by Varouchakis & Hristopulos (2013b). Then for the optimal spatial management of the available dataset the ArcGis software was used to apply interpolation and mapping. The default settings of the Geostatistical Analyst tool regarding the semi-variogram determination were met during the calculation of the optimal semi-variogram in Matlab platform. The latter was necessary for the proper mapping procedure using the ArcGis software. Optimal estimates of semi-variogram model parameters obtained by least squares fit to experimental semi-variogram. The least squares sum for each fitted theoretical semi-variogram model was used as an index of optimal fitting (Varouchakis & Hristopulos, 2013b). The complicated cross-semivariogram for the CoK method was estimated only with the Geostatistical Analyst tool of the ArcGis software and the optimal fitted theoretical model was selected using the provided estimation measures.

Anisotropy was investigated in the Matlab platform by comparing directional semi-variograms in the four main geographical directions (Goovaerts, 1997; Varouchakis & Hristopulos, 2013a) using an angle tolerance of 40^{o}. Smaller tolerance values do not permit a sufficient number of data pairs (> 30) at each lag. Hence, no significant

238 difference among the directional semi-variograms of the different variables of the
239 dataset was detected.

240 For the spatial interpolation approach we use the OK and CoK methods in combination
241 with normalizing transformations. The box-cox transformation was applied to the data
242 using code developed in Matlab to predict the optimal transformation parameter
243 (Varouchakis *et al.*, 2012). The parameter value is then used in interpolation procedure
244 of the Geostatistical Analyst tool. Next the semi-variogram or the cross-semivariogram
245 of the transformed data was determined by testing the most commonly used theoretical
246 models like the Exponential, the Gaussian, the Mattern (K-bessel) and the Spherical
247 model. The spherical semi-variogram provided the optimal fit for both OK and CoK
248 methods. Similar studies have also implemented a spherical semi-variogram (King *et*
249 *al.*, 2010; Banerjee & Siciliano, 2012; Correa-Galeote *et al.*, 2013). Then, OK or CoK
250 is used to derive predictions of the transformed field. The predictions are finally back-
251 transformed to obtain estimates in the initial scale.

252 OK and CoK were used to construct prediction maps for each microorganism. CoK
253 method applied using the most significant environmental variables for each
254 microorganism (multiple stepwise regression) to improve the prediction results. The
255 performance of the kriging-based geostatistical models was evaluated by using the
256 leave one out cross validation technique that is usually applied in small datasets (Witten
257 *et al.*, 2011). A series of well-known statistical measures was employed to compare the
258 true and estimated values of the cross-validation procedure, such as the correlation
259 coefficient R, the Root Mean Square Error (RMSE), the Mean Relative Error (MRE)
260 and Analysis of Variance (ANOVA). In addition, the performance of each geostatistical
261 model was supported by the associated kriging variance plots.

Results

Bacteria were the most abundant domain of microorganisms in all sampling points followed by Fungi (Fig. 2). The abundance of Archaea was two to three orders of magnitube lower compared to that of Fungi and Bacteria. With regard to the bacterial phyla, Acidobacteria phylum was the most abundant, followed by Bacteroidetes (Fig. 2). In comparison to other phyla (α-Proteobacteria, β-Proteobacteria, Actinobacteria and Firmicutes), the abundance of Acidobacteria was one to two orders of magnitude higher. Pearson correlation analysis revealed strong positive relationships between microbial groups (Fig. 3, Table S4). Actinobacteria and Bacteroidetes were the most strongly correlated phyla ($r=0.96$, $p<0.001$). Fungi, on the other hand, showed the weakest relationships with the other groups, except from Acidobacteria ($r=0.52$, $p<0.001$). The ratio of fungal 18S rRNA gene copies to bacterial 16S rRNA gene copies, an indicator commonly employed to draw conclusions for the sustainability of agricultural systems, showed negative correlations with all bacterial taxa and Archaea (Table S4).

Multiple regression analysis showed that a proportion of the variance, ranging from 31.10 to 79.65%, in the distribution of the microbial groups investigated in this study could be explained by the environmental variables monitored (Table 1). Partitioning out overall models R^2, emerged TOC content and pH as the most important predictors with the highest contributions in variance explanation. TOC, by itself, could explain a proportion of variance ranging from 11.7% to 74.8%, while the corresponding proportion of pH varied from 7.36% to 37.14%. Geographical distance also explained a significant proportion (9.49% to 67.48%) of the total variance in some microbial groups (e.g. Firmicutes; Table 1). In accordance, Pearson analysis highlighted TOC and

pH as the most significant variables ($p < 0.001$), (Fig. 3, Table S4). Variance of Fungi domain was the least explained by environmental variables measured.

Complementarily, in order to test if there is a unique microbial group that could improve model prediction, the abundance of microbial group was included into the models. Models predictions were significantly improved, for instance for α-Proteobacteria, 87% of the variation was explained by the models including data on microbial abundance compared to 68% when modeling was solely based on environmental variables (Table S5), but most of the environmental variables were excluded from the models probably due to the strong correlations between microbial groups.

ANOVA of microbial groups and land uses revealed that the abundance of Acidobacteria, Actinobacteria, Bacteroidetes, and Firmicutes was statistically higher at natural ecosystems compared to agricultural lands (Table S6). In addition, the abundance of all microbial groups, except those of α-Proteobacteria and Fungi, increased with elevation, especially at elevations higher than 400 m. It must be noted, however, that land use changes with elevation meaning that natural ecosystems occur mainly at elevations higher than 300 m.

Semi-variogram modeling of microbial groups revealed strong spatial patterns, with autocorrelation length ranging between 309 and 2.244 m. In this study CoK method was adopted to create prediction maps for microbial distribution since it produced more accurate predictions compared to OK with a lower relative mean error, in some cases up to 9%, for all microbial groups except Fungi. In addition, the R^2 between measured and predicted values ranged from 0.306 to 0.575 (Table S7).

Regard to the distribution maps, Actinobacteria, β-Proteobacteria, α-Proteobacteria, Bacteroidetes, and Archaea (Fig. 5b, S1a, 5a, 5d, 4b,) followed more or less the same

spatial pattern, with a spotty and rough separation between areas. The lowest abundances occurred mostly at the west and east sides of the map, while the highest at north-east and south-west. Firmicutes and total Bacteria 16S rRNA gene copies distribution (Fig. S1b, 4a) followed similar spatial patterns with the previous groups, but the areas were smoother and better separated. Acidobacteria and Fungi abundance (Fig. 5c, 4c) followed a more clear distribution and abundance increased with elevation from north to south direction. For all microbial groups, except Acidobacteria and Fungi, the west side of the map represents low abundances and follows exactly the pH and TOC distribution pattern (Fig. S2). At the corresponding part low values of pH and TOC were measured. Fungi to Bacteria ratio distribution map (Fig. 4d) followed the opposite pattern from all the above, and it resembled more the distribution pattern of Fungi.

Discussion

Understanding the drivers regulating the structure of soil microbial communities, their function and their activity comprise important challenges in the modern environmental microbiology. Pertinent information, even at high phylogenetic levels, has been useful to assign taxa with (aggregate) specific functions (Fierer *et al.*, 2007; Philippot *et al.*, 2009; Wessén *et al.*, 2010), to associate broad microbial groups with certain ecosystem services (Six *et al.*, 2006; Averill *et al.*, 2014) and to provide support for a new generation of biogeochemical models explicitly addressing microbial structure (Waring *et al.*, 2013).

Employing qPCR analyses and advanced statistical modeling at the scale of a small-sized Mediterranean watershed, we provide insights on the influence of environmental variables, land use, biochemical activities and microbial interactions on soil microbial

community, which were explored at the domain and phylum level. The ratio of archaeal to bacterial 16S rRNA genes averaged to approximately 0.02, a proportion quite similar to that of archaeal sequences recovered from soils of various ecosystem types (Bates *et al.*, 2011). In this study soil C:N ratio was also the only factor consistently correlated with the relative abundance of Archaea, being higher in soils with lower C:N ratios. An opposite relation however, was identified in our study, h Archaea abundance was positively correlated to C:N ratio (Fig. 3; Table S4). Positive correlations were also set with pH (Pereira e Silva *et al.*, 2012), TOC, NMR and TN as well as with urease and phenol oxidase activity providing evidence for a significant role of Archaea in C and N cycling in Mediterranean ecosystems. Variation analysis showed that TOC, pH and PNR accounted for 55% of the variance observed in Achaea distribution of (Table 1), with the latter predictor displaying a negative regression coefficient. Although negative correlations have been reported between Thaumarchaeota, the most abundant soil archaeal phylum (Thamdrup, 2012), and soil NO_3^--N or NH_4^+-N status (Bates *et al.*, 2011; Pereira e Silva *et al.*, 2012) this was not the case for Koiliaris CZO (Fig. 3; Table S4). Given the correlative nature of the methodology (Ray-Mukherjee *et al.*, 2014), this finding most probably implies an indirect effect of PNR on microorganisms abundance.

The abundance of 18S rRNA genes of Fungi remained constantly lower to that of bacterial 16S rRNA genes across the Koiliaris CZO, even at fields dominated by natural vegetation and/or not-subjected to tillage. Fungal-dominated soils have been associated with better structure (Rillig & Mummey, 2006) and C sequestration (Six *et al.*, 2006) which were not the case for Koiliaris CZO. This pattern may reflect the intense anthropogenic influence that the watershed is subjected, mainly tilling, fertilization and overgrazing (Banwart *et al.*, 2011; Moraetis *et al.*, 2014). Nitrogen addition (Boyle *et al.*, 2008), low soil C inputs (Wang *et al.*, 2014) and overgrazing (Lopez-Sangil *et al.*,

2011) have been linked to a decline in Fungi abundance in soils. Indeed, abundance of Fungi was positively correlated to SOC (Fig. 3). By contrast, no correlation with phenol oxidase was set, as it had been hypothesized. This finding may be due to the fact that the turnover of (poly)phenols is regulated by a specific group of Fungi, the Basidiomycota (Theuerl & Buscot, 2010). In line with other studies (Zinger *et al.*, 2011; Pereira e Silva *et al.*, 2012) the F:B ratio was greater in natural ecosystems compared to agricultural fields (0.29 vs 0.01). Overall, Fungi were weakly linked to the biogeochemical parameters monitored in this study compared to other microbial taxa and only a low proportion of the observed variance could be explained by the co-kriging model. SOM content was the only significant predictor of distribution of Fungi in line with the saprophytic lifestyle of most of Fungi (Boer *et al.*, 2005). Similarly to our findings, Zinger *et al.* (2011) found that only 26% of the fungal variation could be explained in an Alpine landscape when environmental conditions and plant species composition were taken into account. In that study, Fungal beta-diversity was mainly related to SOM, while geographic distance did not account for community changes. The underlying reasons of unexplained variance remain obscure, but they might be related to complex interactions with vegetation, environmental variables (Zinger *et al.*, 2011), the relative narrow range of pH variation, and SOM composition (Pereira e Silva *et al.*, 2012).

The relative abundance of bacterial phyla followed a pattern that diverged substantially from those reported up to date in various soil ecosystems (Philippot *et al.*, 2009; Wessén *et al.*, 2010) including arid ones (Fierer *et al.*, 2005). The most remarkable deviations regarded α-Proteobacteria, Acidobacteria, and Actinobacteria. Acidobacteria were the most abundant bacterial phylum in Koiliaris CZO comprising on average the 28% of bacterial community in line with results from a desert soil (Fierer *et al.*, 2005) as well

384 as from forest soils and croplands at the Brazilian Amazon (Navarrete *et al.*, 2013).
385 Pasternak *et al.* (2013), reported also low abundance of Actinobacteria (0.0013%) and
386 α-Proteobacteria (0.02%) in Mediterranean soils, although considerably lower to that
387 observed in this study. The reasons contributing to this variability of the relative
388 abundance of bacterial phyla across studies remain obscure. It should be underlined that
389 all the pre-mentioned studies have used the same primers, except from the study of
390 Pasternak *et al.* (2013), to amplify bacterial phyla. Although this does not preclude
391 additional methodological bias (e.g. DNA extraction method), it seems more probable
392 that environmental factors cause this variability, stressing the need for more studies to
393 elucidate their influence on the shaping of soil microbial communities.

394 Distribution of bacterial phyla across the Koiliaris CZO was correlated with some
395 geochemical parameters including TOC, TN, C:N ratio, pH, and soil texture (Fig. 5,
396 Table S4) in accordance to other studies (Fierer *et al.*, 2007; Philippot *et al.*, 2009;
397 Nemergut *et al.*, 2011; Navarrete *et al.*, 2013). With regard to the biochemical
398 parameters, urease activity was the only parameter consistently correlated with all
399 bacterial taxa. This finding may suggest the widespread distribution of genes encoding
400 for urease and the great importance of urea in the cycling of C and N in the studied
401 watershed. Compared to other studies (Meyer *et al.*, 2013; Rodrigues *et al.*, 2013), land
402 use had only a slight influence on microbial community structure (Table S6), despite
403 the shifts in the soil management practices applied and the composition of liter entering
404 the soil (olive trees, citrus, natural ecosystems). It should be noted however, that the
405 influence of land use on soil microbial communities may have been confounded by the
406 effect of elevation which followed a quite similar pattern (Table S6), since natural
407 ecosystems in Koiliaris CZO occur mainly at elevations higher than 300 m (Fig. 1 and
408 Fig. S1). These findings may suggest that abiotic factors (climate, geochemistry) had

409 the dominant influence on soil microbial community structure in Koiliaris CZO and
410 this influence may have further exacerbated by the relatively low availability of
411 organic-C.

412 Strong correlations were also set between and within bacterial phyla and domains (Fig.
413 5, Table S4). Although qPCR analysis of bacterial phyla has been found to suffer from
414 caveats, for instance d-Proteobacteria 16S rDNA sequences are also amplified by the
415 qPCR assay of a-Proteobacteria and similarly, Actinobacteria assay amplifies some
416 Verrucomicrobial sequences (Fierer *et al.*, 2005), the low proportion of non-targeted
417 sequences is not expected to have exerted a strong effect on the correlations obtained.
418 These relationships may imply to some extent syntrophic partnerships, but most
419 probably they have been arisen by variations in environmental factors and resources
420 availability and suggest sharing of similar ecological niches by taxonomicaly distinct
421 microorganisms displaying functional redundancy and/or similarity. Network analysis
422 also revealed co-occurrence patterns and non-random association of soil microbial
423 communities implying that these patterns might have been derived from taxa sharing
424 similar ecological niches and did not necessarily imply direct symbioses (Barberan *et*
425 *al.*, 2012).

426 Habitat distribution models have been employed to understand and predict the
427 distribution of microorganisms at various taxonomic levels (King *et al.*, 2010; Bru *et*
428 *al.*, 2011). The prediction power of the models set in Koiliaris CZO, based solely on
429 environmental parameters, differed among the microbial domains and bacterial phyla
430 (Table 1). The TOC, C:N, and pH were identified as the most important chemical
431 predictors. In contrast to previous studies (King *et al.*, 2010), biochemical parameters
432 had a low contribution in explaining microbial variance and only PNR was consistently
433 included as a predictor of some bacterial taxa (Actinobacteria, Firmicutes, Bacterioides)

distribution (Table 1). To the best of our knowledge relations among bacterial taxa and PNR have not studied so far. The negative regression coefficients of the models among bacterial taxa and PNR may suggest competition among microbial taxa for the available soil NH_4^+-N or may have their origin on the strong relationships among microbial taxa as it is indicated in Table 2. However, since studies employing multiple regression are purely correlative, the predictors do not necessarily imply cause-effect relationships. The geographic distance, as expressed by the spatial vectors resulted from PCNM analysis, explained a large proportion of the variation and for same cases (β-Proteobacteria and Fungi to Bacteria ratio) higher than any other individual factor. Moreover semi-variograms revealed strong autocorrelation for microbial groups and co-kriging cross validation revealed high prediction power. These findings suggest that geographic distance is an indicator of bacterial dispersal across Koiliaris CZO. When microbial abundance was incorporated into the models (Table S6) their prediction power was significantly improved, but most of the environmental variables were excluded. Contrary to our expectations, there is not a unique or certain microbial groups that could improve the prediction power of the models and which could be potentially used as bioindicator(s) for predicting microbial community abundance at the domain or phylum level.

Conclusions

In this study we investigate the abundance of soil microorganisms at the domain and the bacterial level across the Koiliaris CZO. This approach allowed us to obtain insights on the environmental drivers regulating the spatial distribution of microorganisms. Multiple regression analysis showed that a percentage from 31% to 79% of the spatial variation in microbial taxa abundance could be explained by the environmental variables measured, while TOC, C:N ratio, pH, PNR and geographic distance were

identified as the most important drivers. Strong correlations among microbial taxa suggesting syntrophic partnerships and/or sharing of similar ecological niches but additional research is needed to shed light on these relationships. Inclusion of microbial taxa abundance in geostatistical models improved strongly their prediction power resulting in variance explanation from 36 to 94%. Our findings contribute on the understanding of environmental factors controlling the abundance and distribution of dominant soil microorganisms at large scale, as well as to define the importance of that influence.

Acknowledgements

The authors acknowledge the funding support from the European Commission FP 7 Collaborative Project Soil Transformations in European Catchments (SoilTrEC) (Grant Agreement No. 244118). We would like to thank Prof J. Bloem and Prof P. de Ruiter for their contributions on the planning of the study.

References

Allison SD, Wallenstein MD & Bradford MA (2010) Soil-carbon response to warming dependent on microbial physiology. *Nature Geosci* **3**: 336-340.

Averill C, Turner BL & Finzi AC (2014) Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature* **advance online publication**.

Baldrian P, Kolarik M, Stursova M, *et al.* (2012) Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J* **6**: 248-258.

Banerjee S & Siciliano SD (2012) Factors Driving Potential Ammonia Oxidation in Canadian Arctic Ecosystems: Does Spatial Scale Matter? *Appl Environ Microbiol* **78**: 346-353.

Banwart S, Bernasconi SM, Bloem J, *et al.* (2011) Soil Processes and Functions in Critical Zone Observatories: Hypotheses and Experimental Design. *Vadose Zone Journal* **10**: 974.

Barberan A, Bates ST, Casamayor EO & Fierer N (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**: 343-351.

Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R & Fierer N (2011) Examining the global distribution of dominant archaeal populations in soil. *ISME J* **5**: 908-917.

Boer Wd, Folman LB, Summerbell RC & Boddy L (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol Rev* **29**: 795-811.

Borcard D & Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model* **153**: 51-68.

Bouyoucos GJ (1962) Hydrometer Method Improved for Making Particle Size Analyses of Soils1. *Agron J* **54**: 464-465.

Boyle SA, Yarwood RR, Bottomley PJ & Myrold DD (2008) Bacterial and fungal contributions to soil nitrogen cycling under Douglas fir and red alder at two sites in Oregon. *Soil Biol Biochem* **40**: 443-451.

Bru D, Ramette A, Saby NP, Dequiedt S, Ranjard L, Jolivet C, Arrouays D & Philippot L (2011) Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *ISME J* **5**: 532-542.

Correa-Galeote D, Marco DE, Tortosa G, Bru D, Philippot L & Bedmar EJ (2013) Spatial distribution of N-cycling microbial communities showed complex patterns in constructed wetland sediments. *FEMS Microbiol Ecol* **83**: 340-351.

Deutsch CV & Journel AG (1992) GSLIB: Geostatistical Software Library and User's Guide. *Oxford University Press, New York, pp 380*.

Fierer N, Bradford MA & Jackson RB (2007) Toward an ecological classification of soil bacteria. *Ecology* **88**: 1354-1364.

Fierer N, Jackson JA, Vilgalys R & Jackson RB (2005) Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays. *Appl Environ Microbiol* **71**: 4117-4120.

Fierer N, Strickland MS, Liptzin D, Bradford MA & Cleveland CC (2009) Global patterns in belowground communities. *Ecol Lett* **12**: 1238-1249.

Fierer N, Ladau J, Clemente JC, Leff JW, Owens SM, Pollard KS, Knight R, Gilbert JA & McCulley RL (2013) Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science* **342**: 621-624.

Goovaerts P (1997) Geostatistics for Natural Resources Evaluation. *Oxford University Press, New York, pp 489*.

Green JL, Bohannan BJM & Whitaker RJ (2008) Microbial Biogeography: From Taxonomy to Traits. *Science* **320**: 1039-1043.

Griffiths RI, Thomson BC, James P, Bell T, Bailey M & Whiteley AS (2011) The bacterial biogeography of British soils. *Environ Microbiol* **13**: 1642-1654.

Hart SC, Nason GE, Myrold DD & Perry DA (1994) Dynamics of Gross Nitrogen Transformations in an Old-Growth Forest: The Carbon Connection. *Ecology* **75**: 880-891.

528 Kandeler E & Gerber H (1988) Short-term assay of soil urease activity using
529 colorimetric determination of ammonium. *Biol Fertil Soils* **6**: 68-72.
530 Kandeler E, Stemmer M & Klimanek E-M (1999) Response of soil microbial biomass,
531 urease and xylanase within particle size fractions to long-term soil management. *Soil*
532 *Biol Biochem* **31**: 261-273.
533 King AJ, Freeman KR, McCormick KF, Lynch RC, Lozupone C, Knight R & Schmidt
534 SK (2010) Biogeography and habitat modelling of high-alpine bacteria. *Nat Commun*
535 **1**: 53.
536 Kitanidis PK (1997) Introduction to Geostatistics. Applications in Hydrogeology.
537 *Cambridge University Press, New York, pp 249*.
538 Lauber CL, Hamady M, Knight R & Fierer N (2009) Pyrosequencing-Based
539 Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the
540 Continental Scale. *Appl Environ Microbiol* **75**: 5111-5120.
541 Legendre P & Legendre L (1998) Numerical Ecology. *ELSEVIER Science BV pp 852*.
542 Li X, Han S, Guo Z, Shao D & Xin L (2010) Changes in soil microbial biomass carbon
543 and enzyme activities under elevated CO2 affect fine root decomposition processes in
544 a Mongolian oak ecosystem. *Soil Biol Biochem* **42**: 1101-1107.
545 Lopez-Sangil L, Rousk J, Wallander H & Casals P (2011) Microbial growth rate
546 measurements reveal that land-use abandonment promotes a fungal dominance of SOM
547 decomposition in grazed Mediterranean ecosystems. *Biol Fertil Soils* **47**: 129-138.
548 Meyer A, Focks A, Radl V, Keil D, Welzl G, Schöning I, Boch S, Marhan S, Kandeler
549 E & Schloter M (2013) Different Land Use Intensities in Grassland Ecosystems Drive
550 Ecology of Microbial Communities Involved in Nitrogen Turnover in Soil. *PLoS One*
551 **8**: e73536.

552 Moraetis D, Stamati F, Kotronakis M, Fragia T, Paranychnianakis N & Nikolaidis NP
553 (2011) Identification of hydrologic and geochemical pathways using high frequency
554 sampling, REE aqueous sampling and soil characterization at Koiliaris Critical Zone
555 Observatory, Crete. *Appl Geochem* **26**: S101-S104.
556 Moraetis D, Paranychianakis NV, Nikolaidis NP, *et al.* (2014) Sediment provenance,
557 soil genesis, and carbon storage in fluvial and manmade terraces at Koiliaris River
558 Critical Zone Observatory. *J Soils Sed* **Under Review**.
559 Navarrete AA, Kuramae EE, de Hollander M, Pijl AS, van Veen JA & Tsai SM (2013)
560 Acidobacterial community responses to agricultural management of soybean in
561 Amazon forest soils. *FEMS Microbiol Ecol* **83**: 607-621.
562 Nemergut DR, Costello EK, Hamady M, *et al.* (2011) Global patterns in the
563 biogeography of bacterial taxa. *Environ Microbiol* **13**: 135-144.
564 Nemergut DR, Schmidt SK, Fukami T, *et al.* (2013) Patterns and Processes of Microbial
565 Community Assembly. *Microbiol Mol Biol Rev* **77**: 342-356.
566 Nielsen UN, Osler GHR, Campbell CD, Burslem DFRP & van der Wal R (2010) The
567 influence of vegetation type, soil properties and precipitation on the composition of soil
568 mite and microbial communities at the landscape scale. *J Biogeogr* **37**: 1317-1328.
569 Nimon K, Oswald F & Roberts JK (2013) yhat: Interpreting Regression Effects. R
570 package version 2.0-0. http://CRAN.R-project.org/package=yhat.
571 Oksanen J, Blanchet FG, Kindt R, *et al.* (2013) vegan: Community Ecology Package.
572 R package version 2.0-9. http://CRAN.R-project.org/package=vegan.
573 Pasternak Z, Al-Ashhab A, Gatica J, Gafny R, Avraham S, Minz D, Gillor O &
574 Jurkevitch E (2013) Spatial and Temporal Biogeography of Soil Microbial
575 Communities in Arid and Semiarid Regions. *PLoS One* **8**.

576 Pereira e Silva MC, Dias ACF, van Elsas JD & Salles JF (2012) Spatial and Temporal
577 Variation of Archaeal, Bacterial and Fungal Communities in Agricultural Soils. *PLoS*
578 *One* **7**: e51554.
579 Peres-Neto PR, Legendre P, Dray S & Borcard D (2006) Variation partitioning of
580 species data matrices: estimation and comparison of fractions. *Ecology* **87**: 2614-2625.
581 Philippot L, Bru D, Saby NPA, Čuhel J, Arrouays D, Šimek M & Hallin S (2009)
582 Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of
583 the 16S rRNA bacterial tree. *Environ Microbiol* **11**: 3096-3104.
584 Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB & Hallin
585 S (2010) The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Micro*
586 **8**: 523-529.
587 R Development Core Team (2013) R: A language and environment for statistical
588 computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-
589 07-0, URL http://www.R-project.org.
590 Ray-Mukherjee J, Nimon K, Mukherjee S, Morris DW, Slotow R & Hamer M (2014)
591 Using commonality analysis in multiple regressions: a tool to decompose regression
592 effects in the face of multicollinearity. *Methods Mol Biol* **5**: 320-328.
593 Rillig MC & Mummey DL (2006) Mycorrhizas and soil structure. *New Phytol* **171**: 41-
594 53.
595 Rodrigues JL, Pellizari VH, Mueller R, *et al.* (2013) Conversion of the Amazon
596 rainforest to agriculture results in biotic homogenization of soil bacterial communities.
597 *Proc Natl Acad Sci U S A* **110**: 988-993.
598 Rousk J, Brookes PC & Bååth E (2010) Investigating the mechanisms for the opposing
599 pH relationships of fungal and bacterial growth in soil. *Soil Biol Biochem* **42**: 926-934.

600 Ruamps LS, Nunan N & Chenu C (2011) Microbial biogeography at the soil pore scale.
601 *Soil Biol Biochem* **43**: 280-286.
602 Serna-Chavez HM, Fierer N & van Bodegom PM (2013) Global drivers and patterns
603 of microbial abundance in soil. *Global Ecol Biogeogr* **22**: 1162-1172.
604 Sinsabaugh RL (2010) Phenol oxidase, peroxidase and organic matter dynamics of soil.
605 *Soil Biol Biochem* **42**: 391-404.
606 Sinsabaugh RL, Gallo ME, Lauber C, Waldrop MP & Zak DR (2005) Extracellular
607 Enzyme Activities and Soil Organic Matter Dynamics for Northern Hardwood Forests
608 receiving Simulated Nitrogen Deposition. *Biogeochemistry* **75**: 201-215.
609 Six J, Frey SD, Thiet RK & Batten KM (2006) Bacterial and Fungal Contributions to
610 Carbon Sequestration in Agroecosystems. *Soil Sci Soc Am J* **70**: 555.
611 Smolders E, Brans K, Coppens F & Merckx R (2001) Potential nitrification rate as a
612 tool for screening toxicity in metal-contaminated soils. *Environ Toxicol Chem* **20**:
613 2469-2474.
614 Strickland MS, Lauber C, Fierer N & Bradford MA (2009) Testing the functional
615 significance of microbial community composition. *Ecology* **90**: 441-451.
616 Thamdrup B (2012) New Pathways and Processes in the Global Nitrogen Cycle. *Annu*
617 *Rev Ecol Evol Syst* **43**: 407-428.
618 Theuerl S & Buscot F (2010) Laccases: Toward disentangling their diversity and
619 functions in relation to soil organic matter cycling. *Biol Fertil Soils* **46**: 215-225.
620 Tsiknia M, Tzanakakis V, Oikonomidis D, Paranychianakis N & Nikolaidis N (2014)
621 Effects of olive mill wastewater on soil carbon and nitrogen cycling. *Appl Microbiol*
622 *Biotechnol* **98**: 2739-2749.

623 Varouchakis EA & Hristopulos DT (2013a) Improvement of groundwater level
624 prediction in sparsely gauged basins using physical laws and local geographic features
625 as auxiliary variables. *Adv Water Resour* **52**: 34-49.
626 Varouchakis EA & Hristopulos DT (2013b) Comparison of stochastic and deterministic
627 methods for mapping groundwater level spatial variability in sparsely monitored basins.
628 *Environ Monit Assess* **185**: 1-19.
629 Varouchakis EA, Hristopulos DT & Karatzas GP (2012) Improving kriging of
630 groundwater level data using nonlinear normalizing transformations-a field application.
631 *Amélioration du krigeage de donńes pízoḿtriques à l'aide de transformations*
632 *normalisatrices non-lińaires Application sur le terrain* **57**: 1404-1419.
633 Venables WN & Ripley BD (2002) Modern Applied Statistics with S. Fourth Edition.
634 Springer, New York.
635 Wang QK, Wang YP, Wang SL, He TX & Liu L (2014) Fresh carbon and nitrogen
636 inputs alter organic carbon mineralization and microbial community in forest deep soil
637 layers. *Soil Biol Biochem* **72**: 145-151.
638 Waring BG, Averill C & Hawkes CV (2013) Differences in fungal and bacterial
639 physiology alter soil carbon and nitrogen cycling: insights from meta-analysis and
640 theoretical models. *Ecol Lett* **16**: 887-894.
641 Wessén E, Hallin S & Philippot L (2010) Differential responses of bacterial and
642 archaeal groups at high taxonomical ranks to soil management. *Soil Biol Biochem* **42**:
643 1759-1765.
644 Wieder WR, Bonan GB & Allison SD (2013) Global soil carbon projections are
645 improved by modelling microbial processes. *Nature Clim Change* **3**: 909-912.

Witten IH, Frank E, Hall MA & . (2011) Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques. *Elsevier Science, San Francisco, pp 664*.

Wright K (2013) corrgram: Plot a correlogram. R package version 1.5. http://CRAN.R-project.org/package=corrgram.

Zinger L, Lejon DPH, Baptist F, Bouasria A, Aubert S, Geremia RA & Choler P (2011) Contrasting Diversity Patterns of Crenarchaeal, Bacterial and Fungal Soil Communities in an Alpine Landscape. *PLoS One* **6**: e19950.

Table 1 Partitioning of the variation of each microbial domain/phylum in the most important explanatory variables.

	Overall model											
	N[1]	ANOVA[2] (F, p)	explained var[3] (%)	Proportion of the total variance unique explanation by each predictor (%)							Co-variation (%)	Unexplained (%)
α-Proteobacteria	6	19.08 ***	68.02	pH (16.78) ***	TOC (6.86) **	C/N (3.64) *	Sp. Dist $_{V2}$ (4.86) *	Sp. Dist $_{V4}$ (6.92) **	Sp. Dist $_{V23}$ (7.21) **		53.75	31.98
β-Proteobacteria	4	11.66 ***	49.8	TOC (11.73) *	Sp. Dist $_{V2}$ (21.94) **	Sp. Dist $_{V5}$ (17.8) **	Sp. Dis$_{V20}$ (11.92) *				36.61	50.2
Actinobacteria	7	29.51 ***	79.65	TOC (26.4) ***	pH (24.51) ***	PNR (6.43) ***	PhO (2.34) *	Sp. Dist $_{Y}$ (10.47) ***	Sp. Dist $_{V3}$ (5.56) **	Sp. Dist $_{V20}$ (3.11) *	21.18	20.35
Acidobacteria	2	26.02 ***	49.51	TOC (74.8) ***	NO3_N (25.01) ***						0.19	50.5
Bacteroidetes	7	25.25 ***	76.9	TOC (25.01) ***	pH (13.65) ***	PNR (5.54) **	Sp. Dist $_{V1}$ (3.21) *	Sp. Dist $_{V2}$ (2.66) *	Sp. Dist $_{V3}$ (7.1) ***	Sp. Dist $_{V10}$ (2.4) *	40.43	23.1
Firmicutes	5	32.57 ***	52.7	TOC (20.84) ***	pH (37.14) ***	PNR (9.25) *	Elev (10.09) *	Sp. Dist $_{V3}$ (10.32) *			12.36	47.3
Total Archaea	5	23.54 ***	68.85	TOC (23.99) ***	pH(22.9) ***	PNR (8.02) **	Sp. Dist $_{V2}$ (3.71) *	Sp. Dist $_{V3}$ (7.84) **			33.54	31.15
Total Fungi	2	12.50 ***	31.08	TOC (98.06) ***	Soil Moist (45.64) **						-43.7	68.92

Total Bacteria	5	28.32 ***	78.95	TOC (40.19) ***	pH (36.6) ***	NO_3_N (7.08) ***	PNR (10.15) ***	Sp. Dist $_{V3}$ (3.87) **	Sp. Dist $_{V20}$ (3.01) *	Sp. Dist $_{V27}$ (2.61) *	-3.57	21.05
Fungi/Bacteria ratio	7	22.25 ***	74.47	pH (17.36) ***	Sp. Dist $_{V2}$ (14.33) ***	Sp. Dist $_{V5}$ (7.03) **	Sp. Dist $_{V9}$ (6.14) **	Sp. Dist $_{V11}$ (2.85) *	Sp. Dist $_{V13}$ (3.5) *	Sp. Dist $_{V17}$ (23.63) ***	25.16	25.53

[1]N: number of explanatory variables in the final model
[2]ANOVA tests the goodness of fit of the model and its significance
[3]Total explained variance from the overall model calculated by adjusting R^2 values, in order to obtain unbiased estimates (Peres-Neto et al., 2006).
Bonferroni –correction was applied to p values to maintain the family-wise error level in multiple testing.
Abbreviations: TOC: total organic carbon; PNR: potential nitrification rate; PhO: phenol oxidase; Elev.: elevation; Sp. Dist.: spatial distance vector

Figure captions

Figure 1 Sampling points across Koiliaris CZO. The orange-colored line indicates the borders of the watershed while the blue one indicates the hydrographic network, the length of which approaches 36 km. Sampling points are grouped into two broad land uses, agricultural lands (yellow circles) and natural ecosystems (green circles).

Figure 2 Variation in the abundance (gene copy numbers per g of dry weight soil) of microbial domains and bacterial phyla across the Koiliaris CZO. The upper and lower boundaries indicate the 75^{th} and the 25^{th} percentile, respectively; the mid-line indicates the median of the distribution; above and below whiskers indicate the 90^{th} and 10^{th} percentiles, respectively; the black asterisks indicate values identified as outliers.

Figure 3 Correlogram representing Pearson's correlation coefficient rank between and among soil properties and microbial community abundances. All parameters, except pH, were transformed before statistical analysis. Information on the transformations applied is included in Table S3. More detailed information on the significance of the correlations as well as on the correlation coefficients can be found in Table S4.

Figure 4 Maps of the abundance of microbial domains in the Koiliaris CZO generated through co-kriging. (a) Bacterial 16S rRNA gene copies, (b) Archaeal 16S rRNA gene copies, (c) Fungal 18S rRNA gene copies, and (d) Bacteria:Fungi ratio. The color scale at the left of the maps indicates the abundance values (gene copies no/g of soil d.w.) except in the case of Bacteria:Fungi ratio which indicates proportion.

Figure 5 Maps of the abundance of bacterial phyla in the Koiliaris CZO generated through co-kriging. (a) α-Proteobacterial 16S rRNA gene copies, (b) Actinobacterial 16S rRNA gene copies and (c) Acidobacterial 16S rRNA gene copies, (d) Bacteroidetes 16S rRNA gene copies. The color scale at the left of the maps indicates the abundance values (gene copies no/g of soil d.w.).

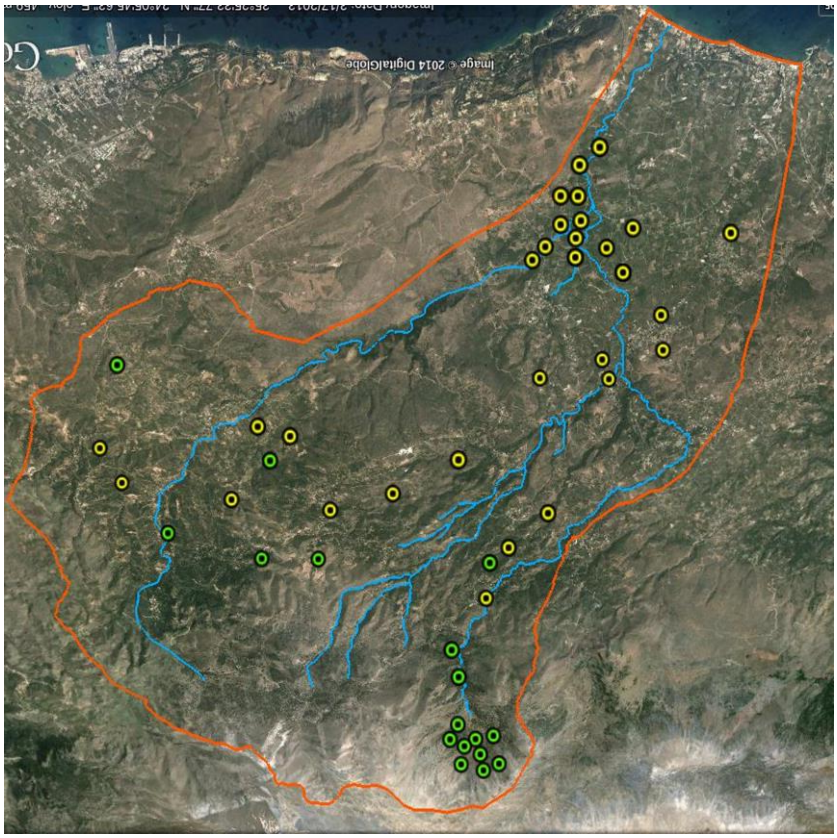


Figure 1: Sampling points across Koiliaris CZO. The orange-colored line indicates the borders of the watershed while the blue one indicates the hydrographic network, the length of which approaches 36 km. Sampling points are grouped into two broad land uses, agricultural lands (yellow circles) and natural ecosystems (green circles).

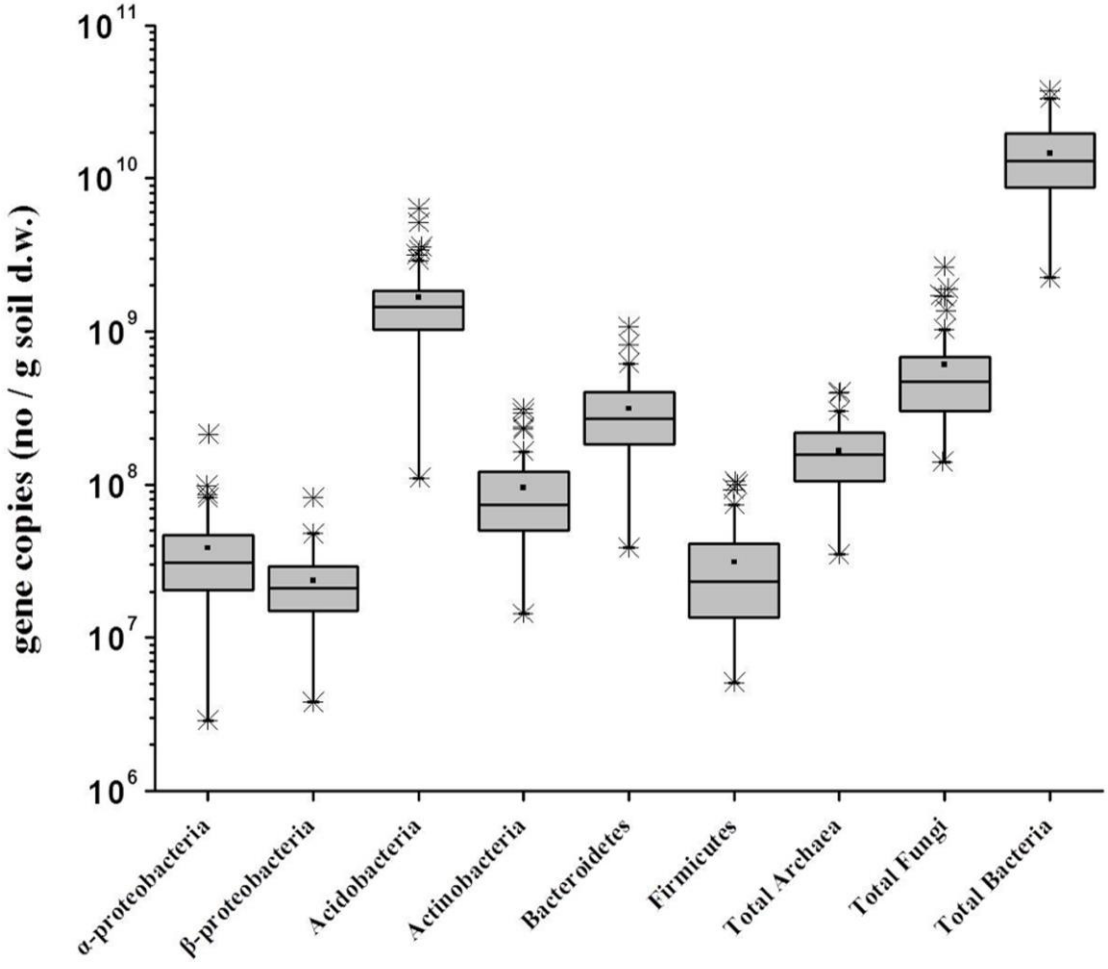


Figure 1: Variation of the abundance 16S/18S rRNA gene copies (no/g soil d.w) of microbial domains and bacterial phyla across the Koiliaris CZO. The upper and lower boundaries indicate the 75th and the 25th percentile, respectively; the mid-line indicates the median of the distribution; above and below whiskers indicate the 90th and 10th percentiles, respectively; the black asterisks indicate values identified as outliers.

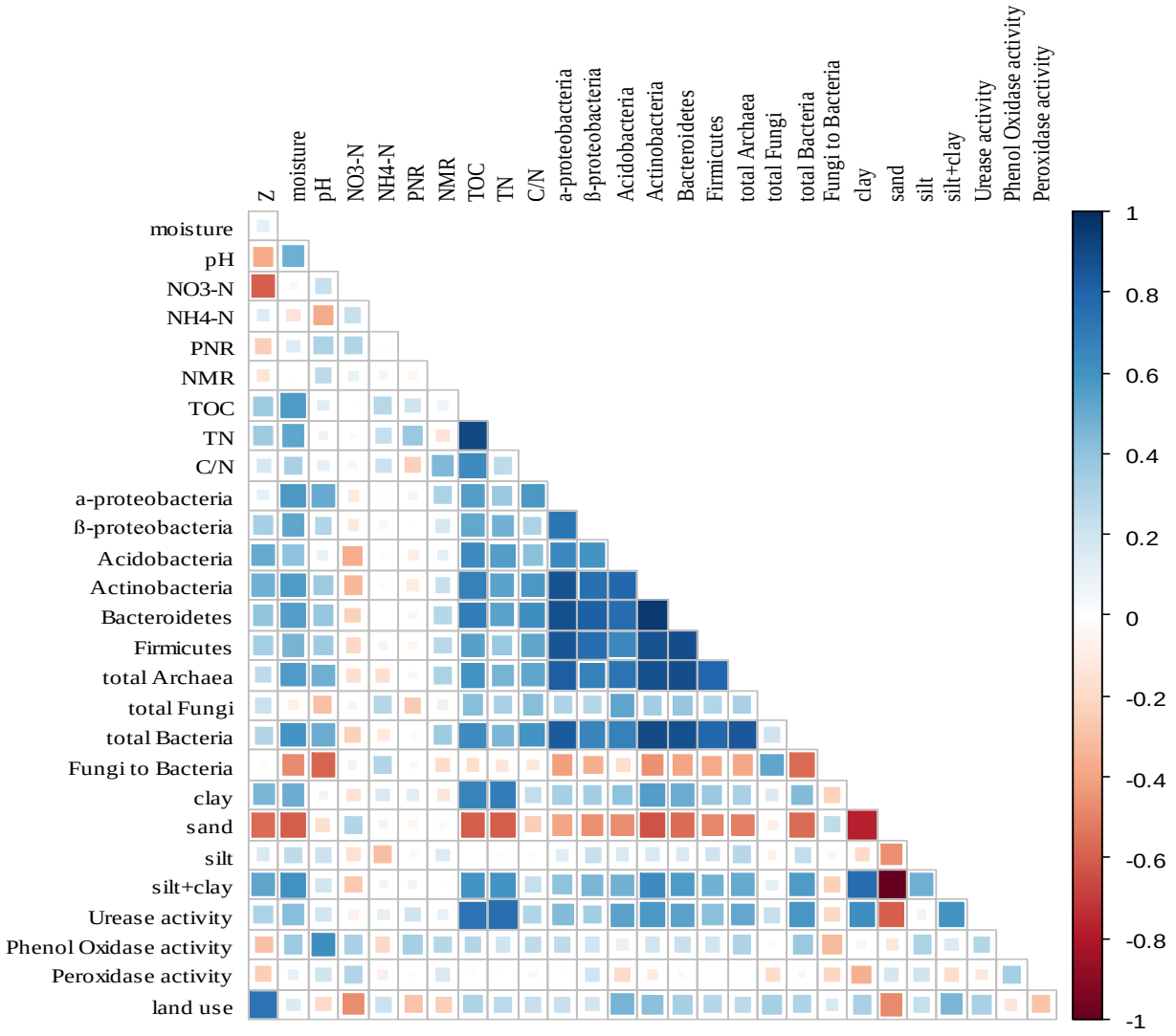


Figure 2: Correlogram representing Pearson's correlation coefficient rank between and among soil properties and microbial community abundances. All parameters, except pH, were transformed before analysis. Information on the transformations applied is included in Table S3. The correlation coefficients and the significance are summarized in Table S4.

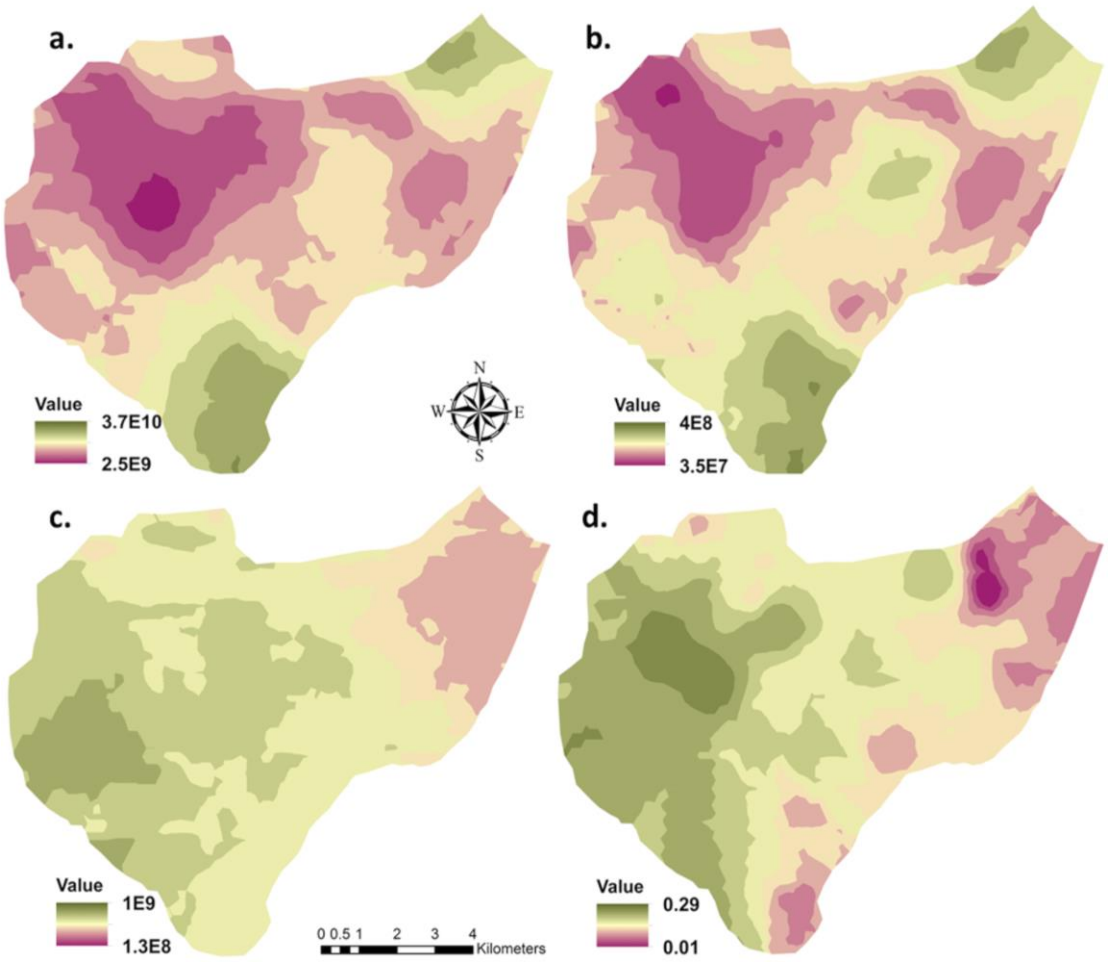
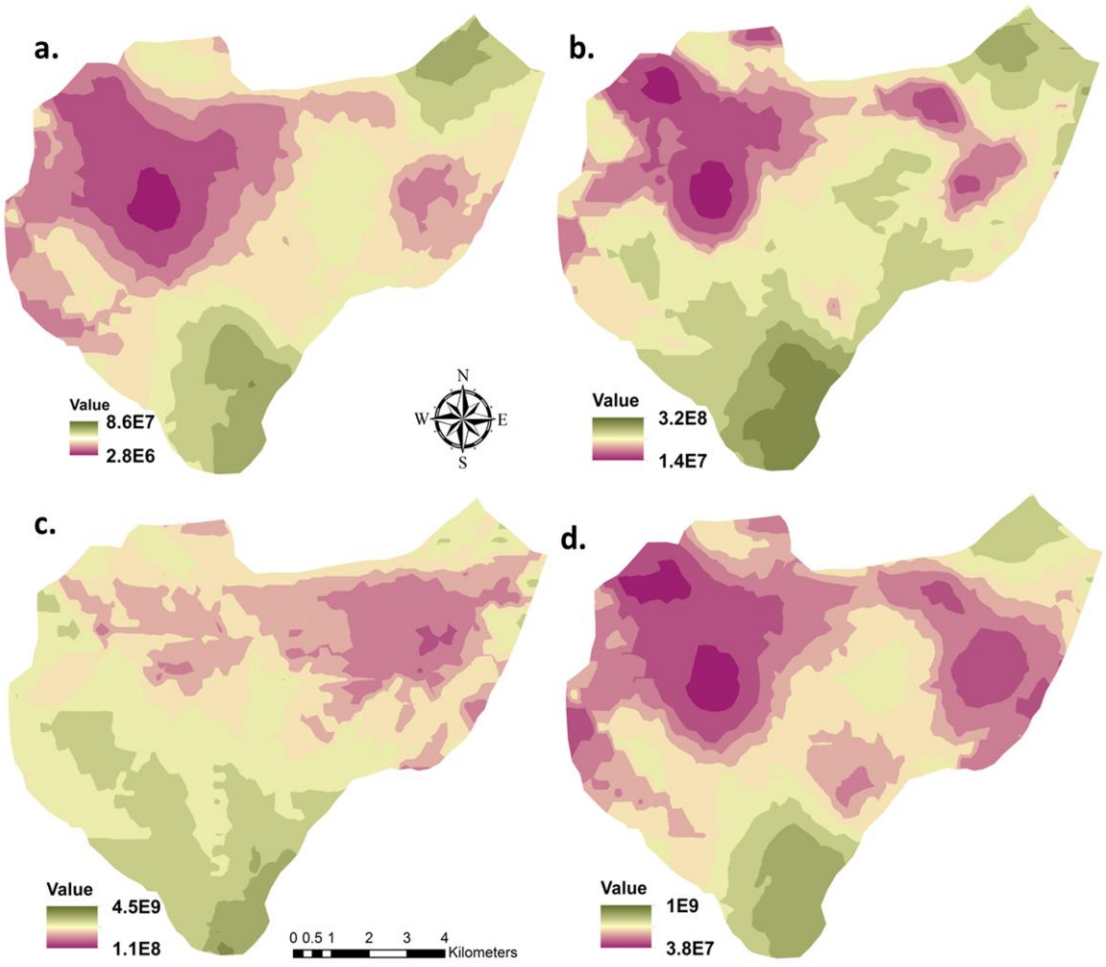


Figure 3: Maps of the abundance of microbial domains in the Koiliaris CZO generated through co-kriging. (a) Bacterial 16S rRNA gene copies, (b) Archaeal 16S rRNA gene copies, (c) Fungal 18S rRNA gene copies, and (d) Bacteria:Fungi ratio. The color scale at the left of the maps indicates the abundance values (gene copies no/g soil d.w.) except in the case of Bacteria:Fungi ratio which indicates proportion.

718



719

720 Figure 4: Maps of the abundance of bacterial phyla in the Koiliaris CZO generated
721 through co-kriging. (a) α-Proteobacterial 16S rRNA gene copies, (b) Actinobacterial
722 16S rRNA gene copies and (c) Acidobacterial 16S rRNA gene copies, (d) Bacteroidetes
723 16S rRNA gene copies. The color scale at the left of the maps indicates the abundance
724 values (gene copies no/g soil d.w.).