

Supplementary Material

The impact of logging and forest conversion to oil palm on soil bacterial communities in Borneo

Larisa Lee-Cruz¹, David P. Edwards^{2,3}, Binu Tripathi¹, Jonathan M. Adams^{1*}

¹ Department of Biological Sciences, Seoul National University, Gwanak, Seoul 151, Republic of Korea.

² Centre for Tropical Environmental and Sustainability Science (TESS) and School of Marine and Tropical Biology, James Cook University, Cairns, Queensland 4878, Australia.

³ Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom.

*Corresponding author

E-mail: jonadams@snu.ac.kr

Tel: +82-2-880-4404

Text S1

Variation Partitioning Analysis

We performed a redundancy analysis (RDA) and variation partitioning (Peres-Neto *et al.* 2006) to see whether soil parameters, land use or spatial distance influence bacterial community composition. We used the Hellinger-transformed OTU abundance as our response variable. Soil parameters included pH, total nitrogen, total carbon and available phosphorous. C/N ratio was highly correlated with total carbon (Spearman $\rho^2 = 0.75$), thus was removed from this analysis. For spatial variables we considered a trend present in our data using the latitude and longitude coordinates of each site midpoint, as well as the eigenfunctions generated by a Principal Components of Neighbour Matrices analysis (PCNM; Borcard & Legendre 2002; Borcard *et al.* 2004).

The detection of spatial structure on any variable depends to a certain extent on the distance between sampling points (i.e. spatial scale; Bellehumeur & Legendre 1998; Borcard & Legendre 2002). In the PCNM analysis the threshold distance that is specified influences greatly the PCNM values generated. Ideally the distance value should be the longest to keep the data connected. In irregular sampling designs such as ours a distance value that retains the more distant pair of points prevents detection of finer spatial structures (Borcard & Legendre 2002). Four of our samples from oil palm plantations were at least 50 km away from the rest of the sampling points, five times farther than the longest distance between two sampling points for the rest of the samples. Although it is possible to add additional coordinates where no sampling was done to reduce the truncation distance, this causes a loss on the orthogonality among the coordinates (Borcard *et al.* 2004). Thus, we decided to leave these four samples out of the analysis to increase the likelihood of detecting different levels of spatial structure on our data.

The PCNM analysis generated six positive PCNM eigenfunctions which were used later in the RDA analysis. Before the variation partitioning analysis, we first performed separate RDA's for the spatial trend (Euclidean distance), the environmental variables and the PCNM eigenfunctions. These three models were significant (all $P < 0.05$). We then used forward selection (Blanchet *et al.* 2008) and removed non-significant variables from each of the three explanatory sets. Latitude, pH and available phosphorous, and two PCNM eigenfunctions were the only significant variables after forward selection. To partition the variation in the bacterial community composition we used these three explanatory sets and a fourth one that contained land use recoded as a dummy variable. Test of significance for each explanatory set was done by 999 permutations. This analysis was performed using the vegan, PCNM and ade4 packages in R version 2.14.1

References

Bellehumeur C, Legendre P (1998). Multiscale sources of variation in ecological variables: modelling spatial dispersion, elaborating sampling designs. *Landscape Ecol* 13:15-25.

Blanchet FG, Legendre P, Borcard D (2008). Forward selection of explanatory variables. *Ecology* 89:2623-2632.

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.

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004). Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85:1826-1832.

Dray S, Dufor AB (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Software*

Legendre P, Borcard D, Blanchet G, Dray S (2010) PCNM: PCNM spatial eigenfunction and principal coordinate analyses. R package version 2.1/r82

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB et al. (2011). *vegan: Community Ecology Package*. R package version 2.0-2 Simpson GL et al.

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87:2614-2625.

Figure S1

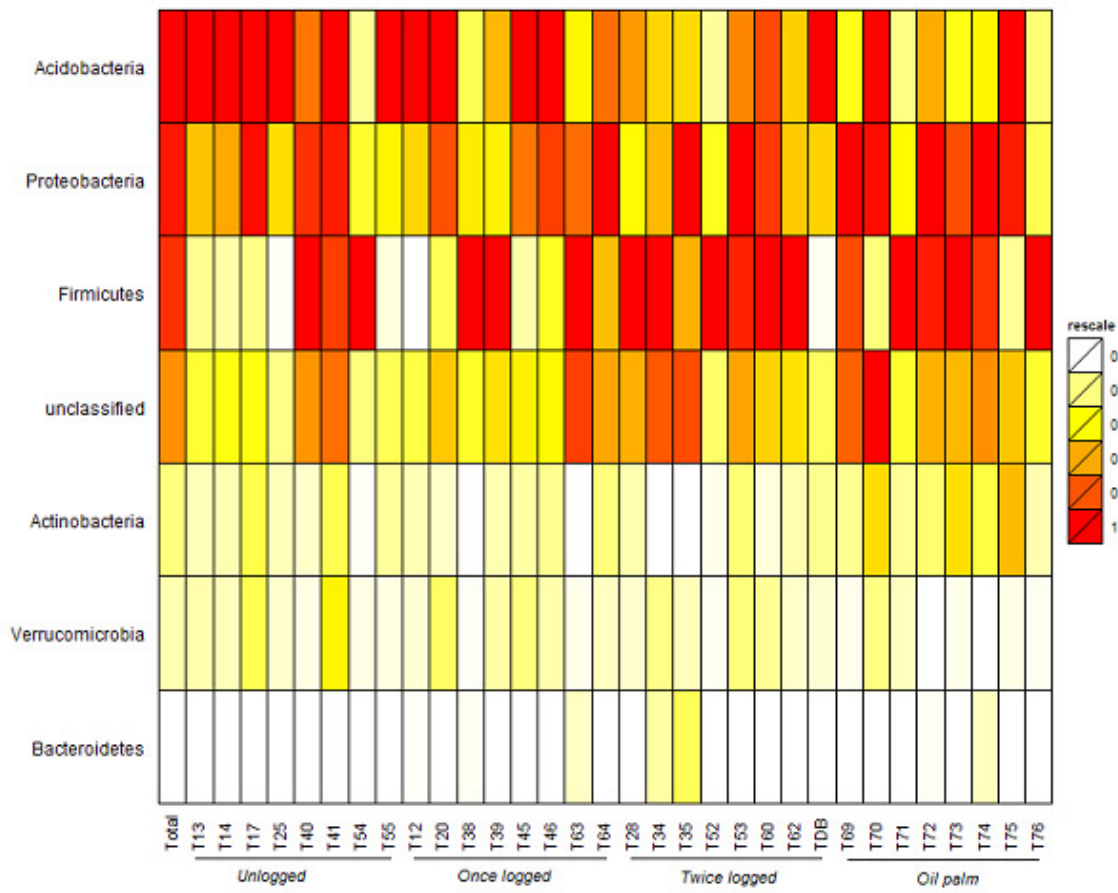


Fig.S1 Relative abundance of the most common phyla in relation to land use. These phyla represented over 94% of the sequences in each group.

Figure S2

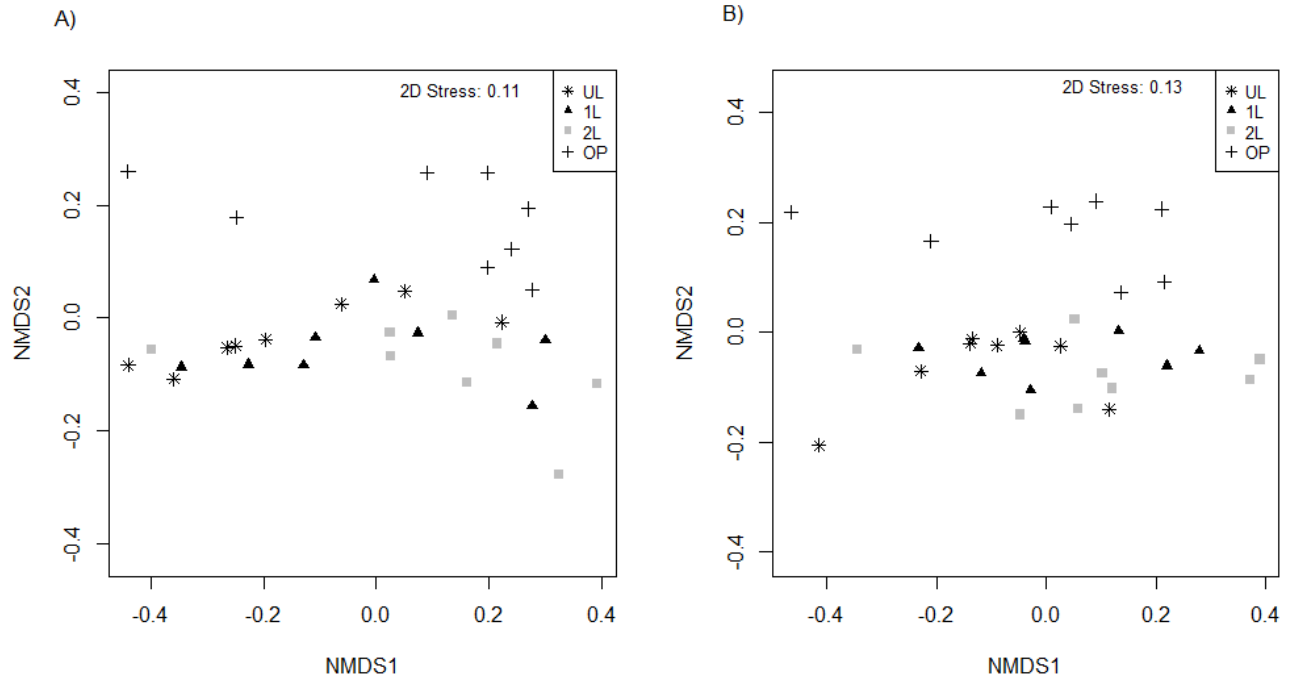


Fig. S2 NMDS of Bray-Curtis (A) and unweighted-Unifrac (B) of bacterial community composition in relation to land use (UL: unlogged, 1L: once-logged, 2L: twice-logged, OP: oil palm).

Figure S3

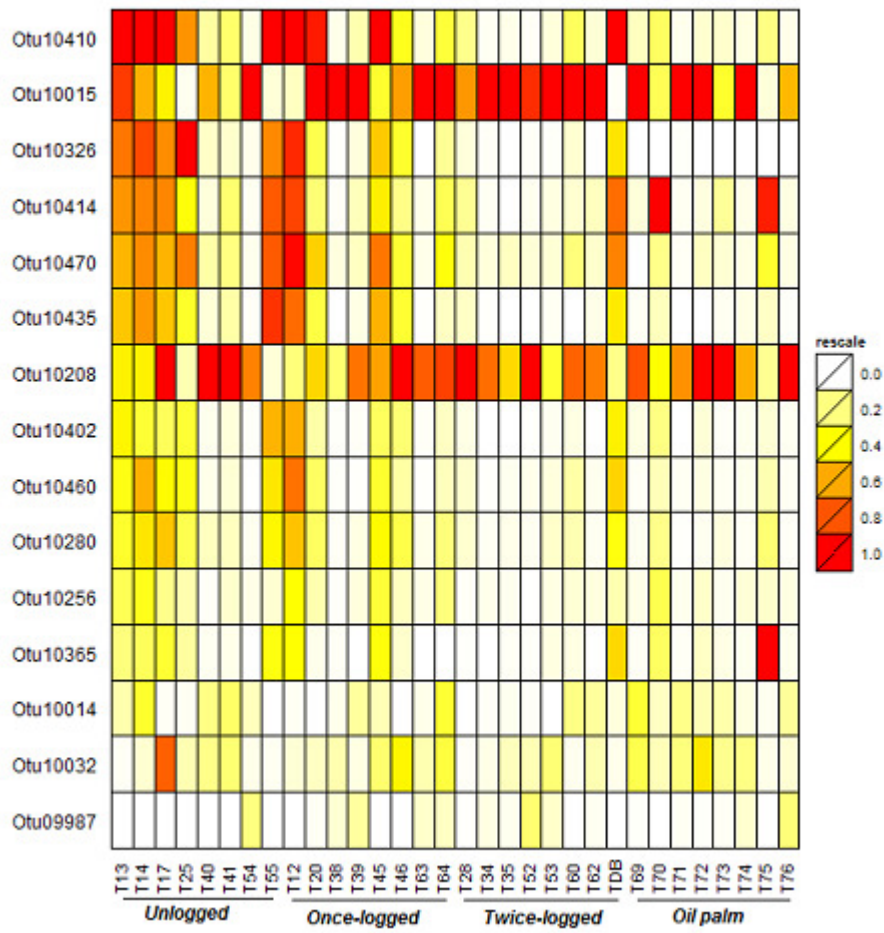


Fig. S3 Relative abundance of the 15 most common OTUs across land use types. Together these OTUs represent 31.7% of all sequences.

Figure S4

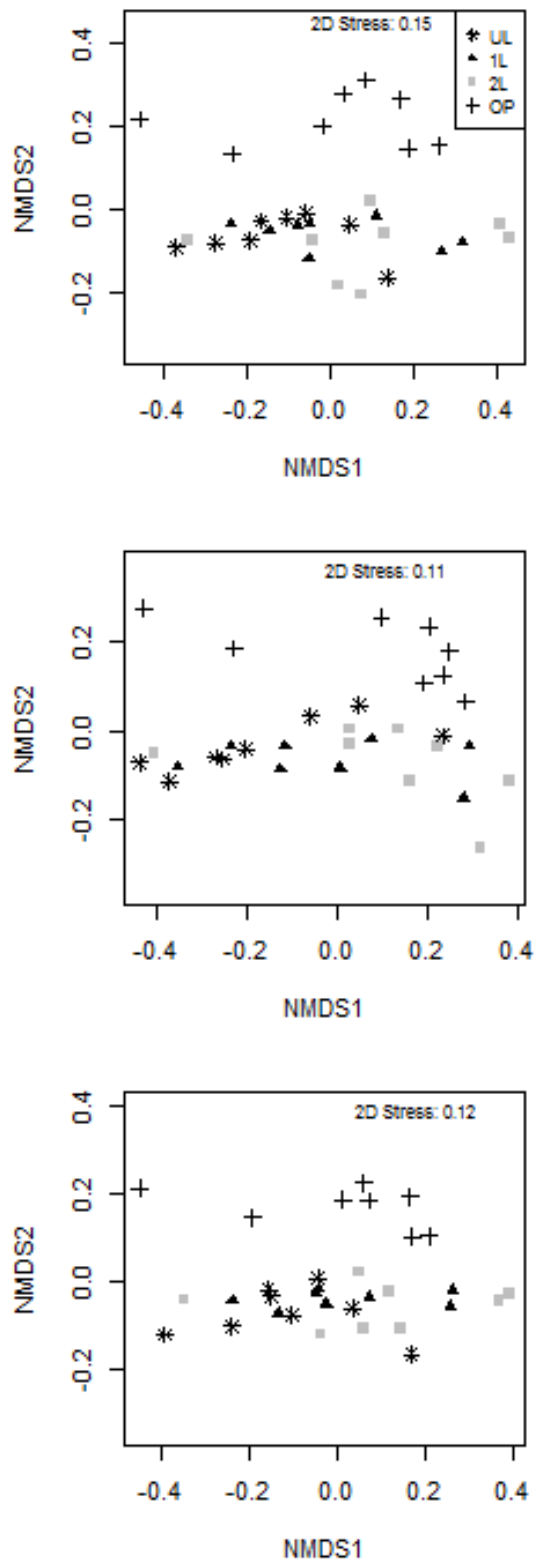


Fig. S4 NMDS of Jaccard (top), Bray-Curtis (middle) and unweighted-Unifrac (bottom) of bacterial community composition in relation to land use (UL: unlogged, 1L: once-logged, 2L: twice-logged, OP: oil palm) with singleton OTUs removed.

Figure S5

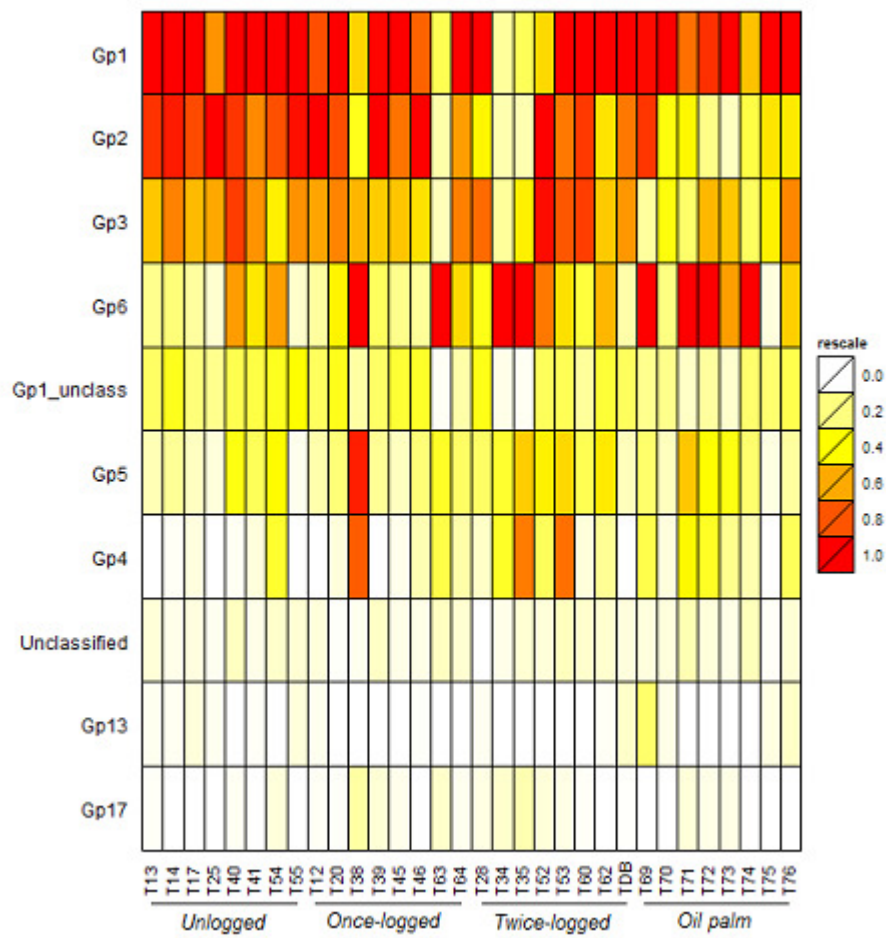


Fig. S5 Relative abundance of the 10 most abundant orders within Acidobacteria in each land use. These orders represent 94.4% of the sequences within Acidobacteria, and 21.4% of all sequences.

Figure S6

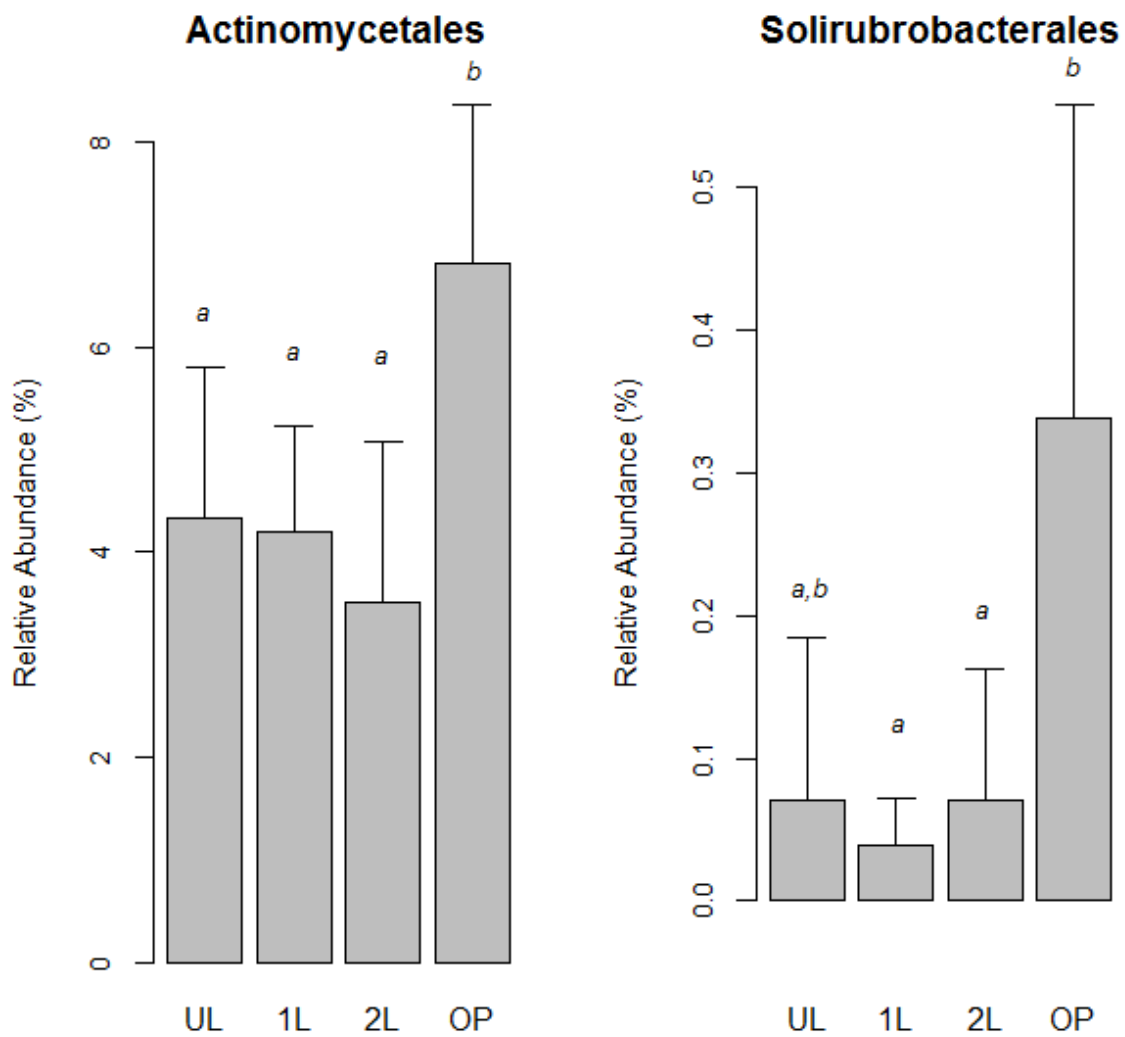


Fig. S6 Relative abundance (Mean \pm SD) in relation to land use (UL: unlogged, 1L: once-logged, 2L: twice-logged, OP: oil palm) for Actinomycetales ($F_{3,28}=8.24$, $P=0.0004$) and Solirubrobacterales (Kruskal-Wallis, $\chi^2 = 13.8$, $P=0.003$). Different letters represent significant differences at $P<0.05$

Figure S7

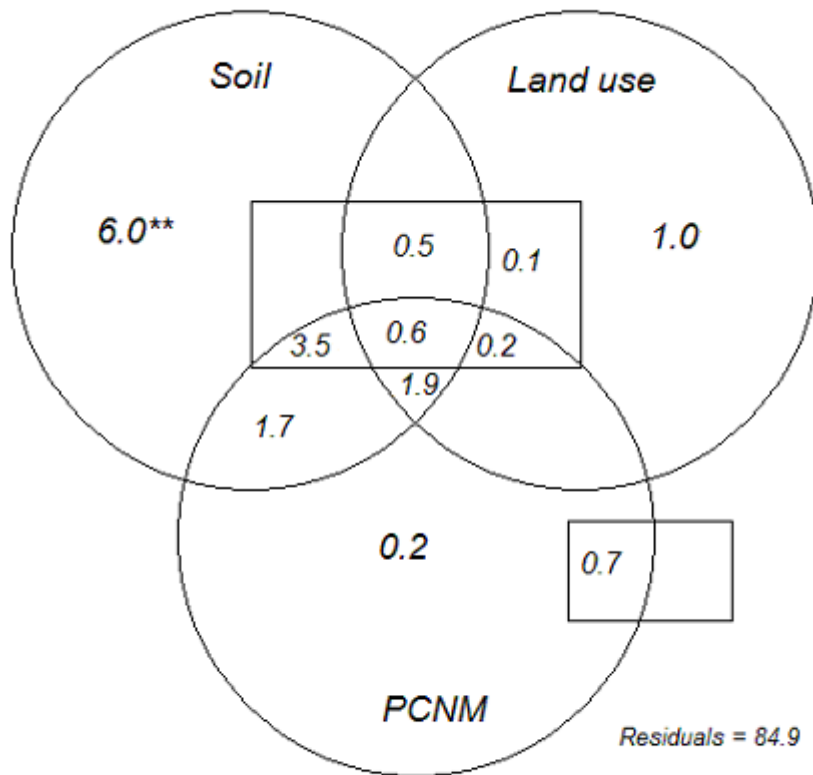


Fig.S7 Percentage of variation explained by soil parameters, land use, PCNM eigenfunctions and spatial trend (shown in rectangles). Unique fractions of each explanatory set were evaluated by 999 permutations. Only the variation explained by soil parameters was significant (**P = 0.005).