Supplementary materials for "Agricultural land use induces broader homogenization of soil microbial functional composition than taxonomic composition in sub-Saharan Africa"

Takamitsu Ohigashi, Yvonne M. Madegwa, George N. Karuku, Keston Njira, Yoshitaka Uchida

Contents:

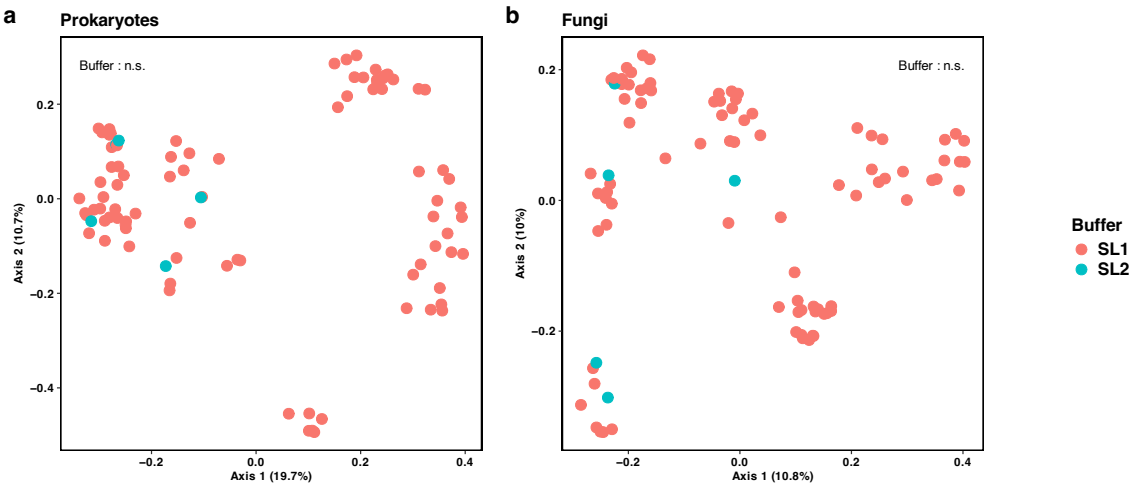

Figure S1. Effect of DNA extraction buffer on microbial community composition. Principal coordinate analysis (PCoA) plots for (**a**) prokaryotes and (**b**) fungi are shown. The p-values from the PERMANOVA accounting for buffer type are indicated as "n.s.", representing no significant difference.

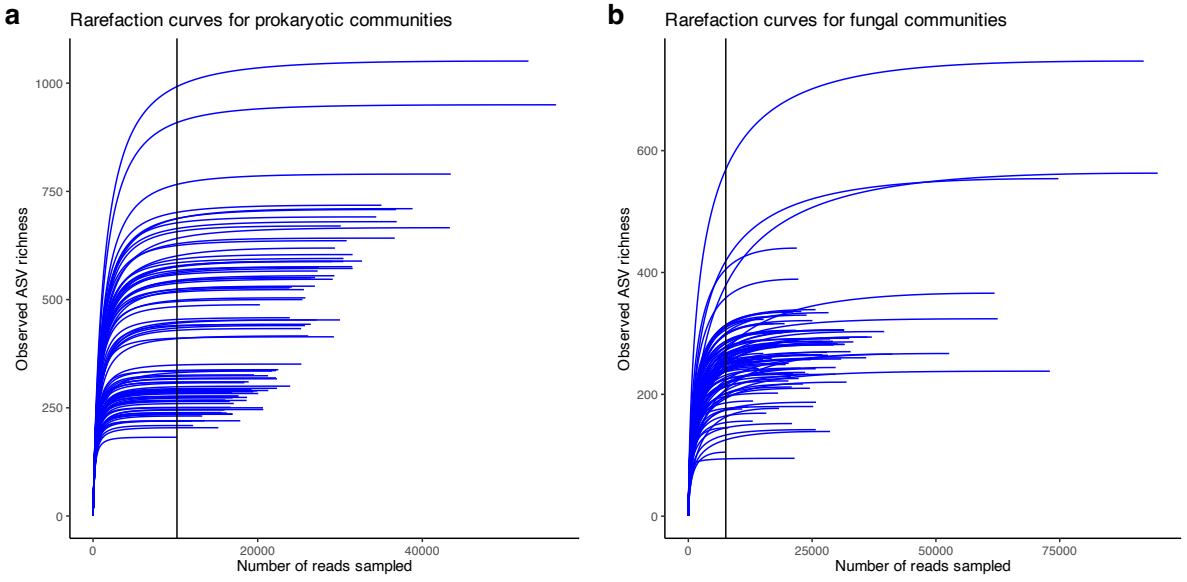

Figure S2. Rarefaction curves for prokaryotic and fungal communities across all samples. Rarefaction curves were generated based on ASV counts from 16S rRNA (**a**) and ITS (**b**) amplicon sequencing data. Blue lines show the number of observed ASVs at each sequencing depth. Black vertical lines represent the minimum read thresholds used for rarefaction (10,193 for prokaryotes and 7,564 for fungi).

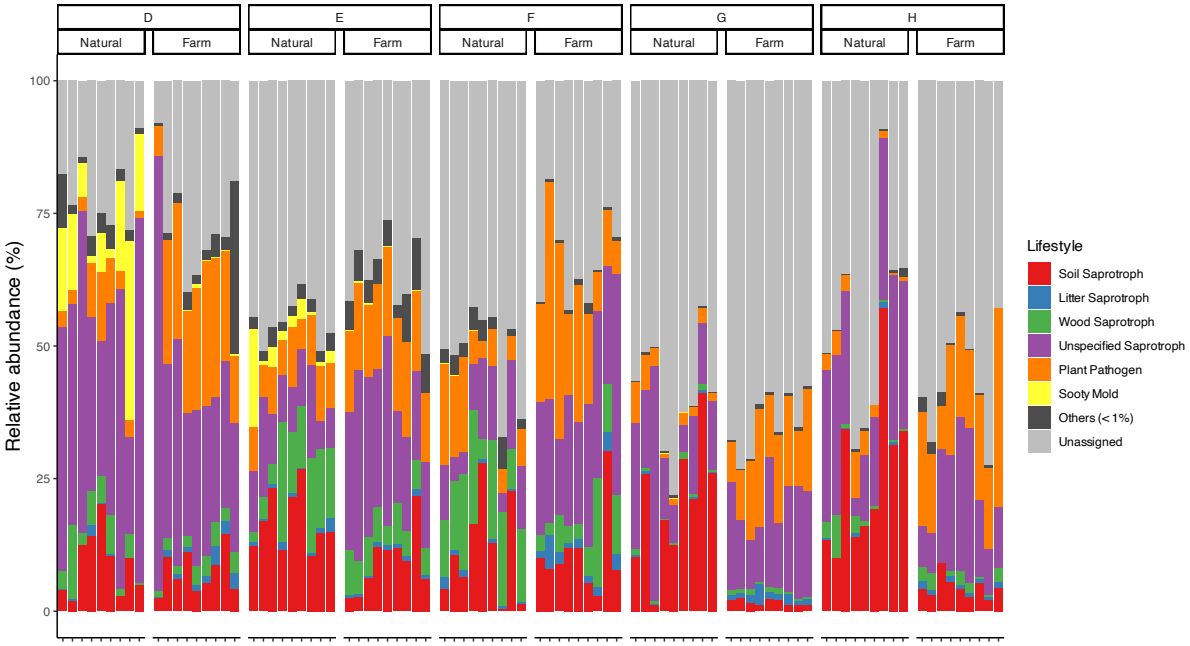

Figure S3. Proportion of fungal ASVs annotated with primary lifestyles based on FungalTraits. The primary lifestyles with mean realtive abundances of > 1% are shown. Based on FungalTraits annotations, 22.9–92.1% of ASVs per sample were assigned a primary lifestyle, with a mean annotation rate of 56.0%.

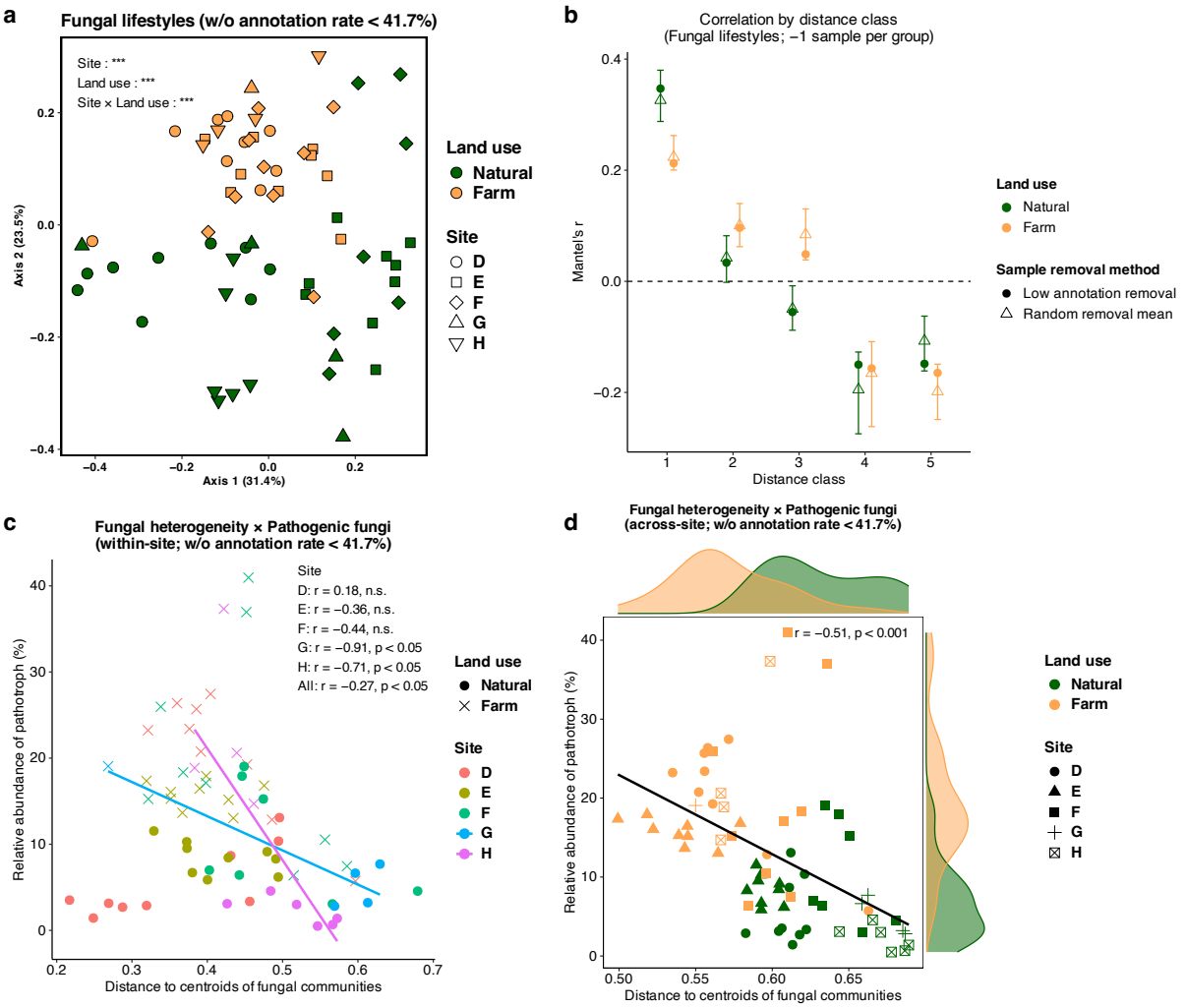

Figure S4. Sensitivity analysis excluding samples with low FungalTraits annotation rates. (**a**) Principal coordinates analysis (PCoA) of fungal functional compositions, based on a reduced dataset excluding samples with annotation rates below the first quartile (n = 67). The overall pattern of functional homogenization among farmland communities remained consistent with the main text results (Figure 2d). (**b**) Mantel correlogram of fungal functional compositions, based on a dataset excluding the sample with the lowest annotation rate from each treatment group. Mantel r values from this reduced dataset (filled circles) are compared with the distribution obtained from 1,000 iterations of randomly excluding one sample per treatment group. Triangles represent the mean r values from the random exclusions, and whiskers indicate their 95% confidence intervals. (**c–d**) Relationships between fungal taxonomic heterogeneity (distance to centroid) and the relative abundance of pathotrophs, based on the reduced dataset (n = 67). While correlations varied by site, the overall negative trend observed in the main analysis (Figure 6a–b) was preserved.

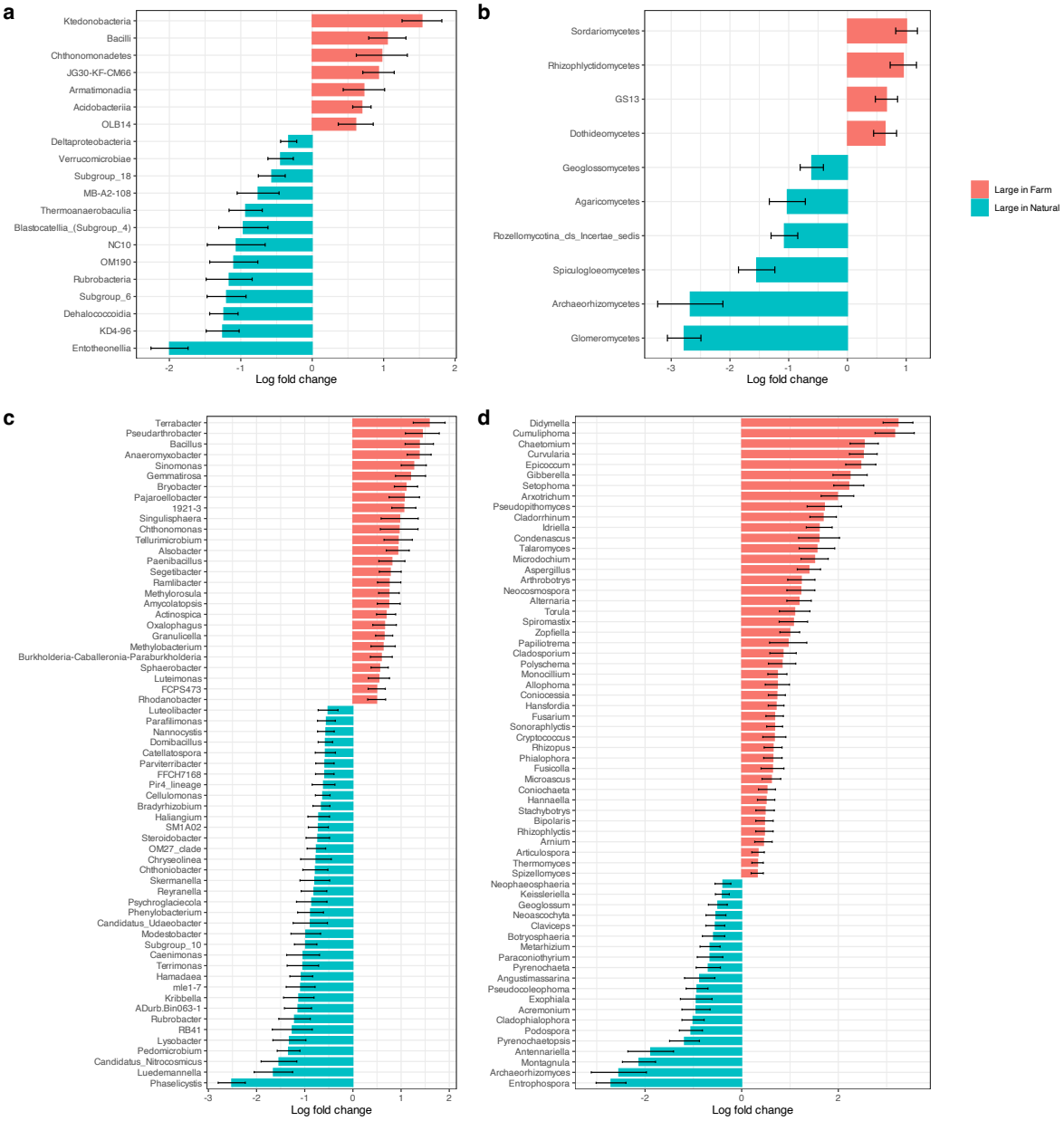

Figure S5. Microbial taxa significantly changed the abundances by land use. Taxa significantly changed their abundances between the land uses, tested by ANCOMBC were shown at class level (**a, b**) and genus level (**c, d**), for prokaryotes (**a, c**) and fungi (**b, d**). Log_{10} fold changes of relative abundances in farmlands compared to natural lands across all sites are calculated. Microbial taxa that showed adjusted p-values < 0.05 are illustrated.

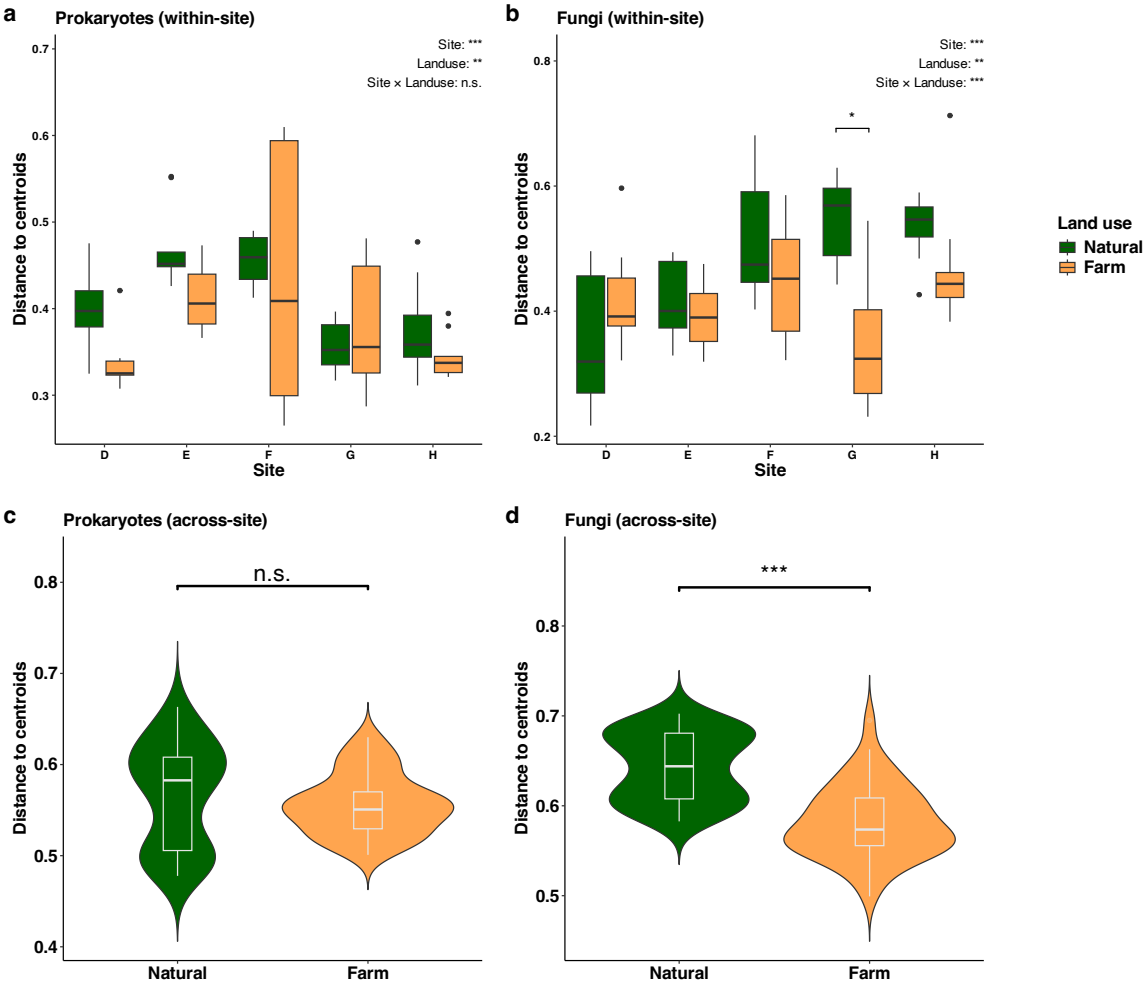

Figure S6. Dissimilarity of soil microbial taxa in natural lands and farmlands. The distance to centroids of the taxonomic compositions in each land use in (**c, d**) within-site and (**e, f**) across-site scales are plotted, for prokaryotes (**c, e**) and fungi (**d, f**). The p-values in the two-way ANOVA on sites, land uses and their interaction, and those in the t-test on land uses are indicated with "*", "**", or "***", representing $p < 0.05$, $p < 0.01$, or $p < 0.001$, respectively. If a significant interaction was found in the two-way ANOVA, pairwise comparisons of estimated marginal means were conducted to assess whether there were significant differences in land use within each site. Significant differences in land use within sites were indicated by a single asterisk "*".

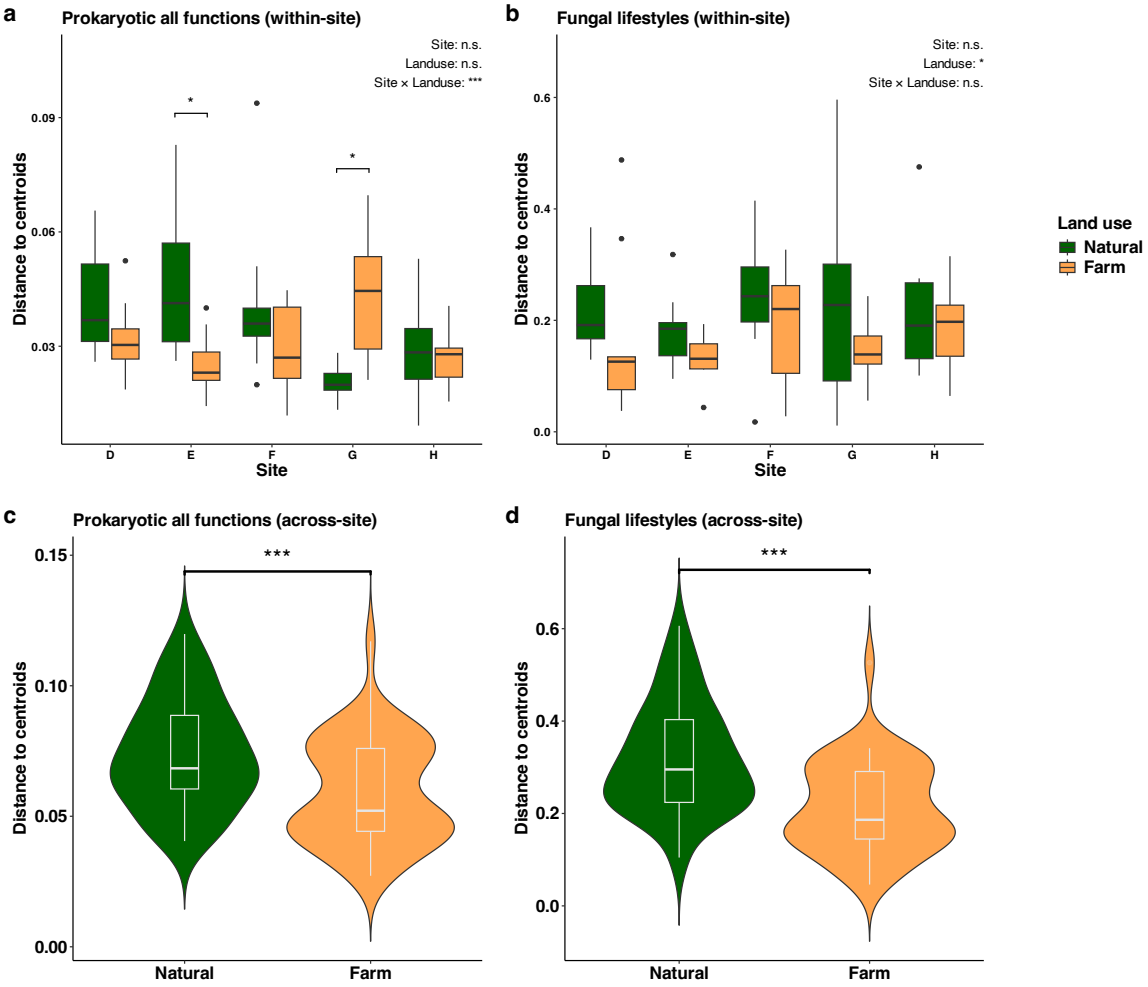

Figure S7. Dissimilarity of soil microbial functions in natural lands and farmlands. The distance to centroids of the functional compositions in each land use in (**c, d**) within-site and (**e, f**) across-site scales are plotted, for prokaryotes (**c, e**) and fungi (**d, f**). The p-values in the two-way ANOVA on sites, land uses and their interaction, and those in the t-test on land uses are indicated with "*", "**", or "***", representing $p < 0.05$, $p < 0.01$, or $p < 0.001$, respectively. If a significant interaction was found in the two-way ANOVA, pairwise comparisons of estimated marginal means were conducted to assess whether there were significant differences in land use within each site. Significant differences in land use within sites were indicated by a single asterisk "*".

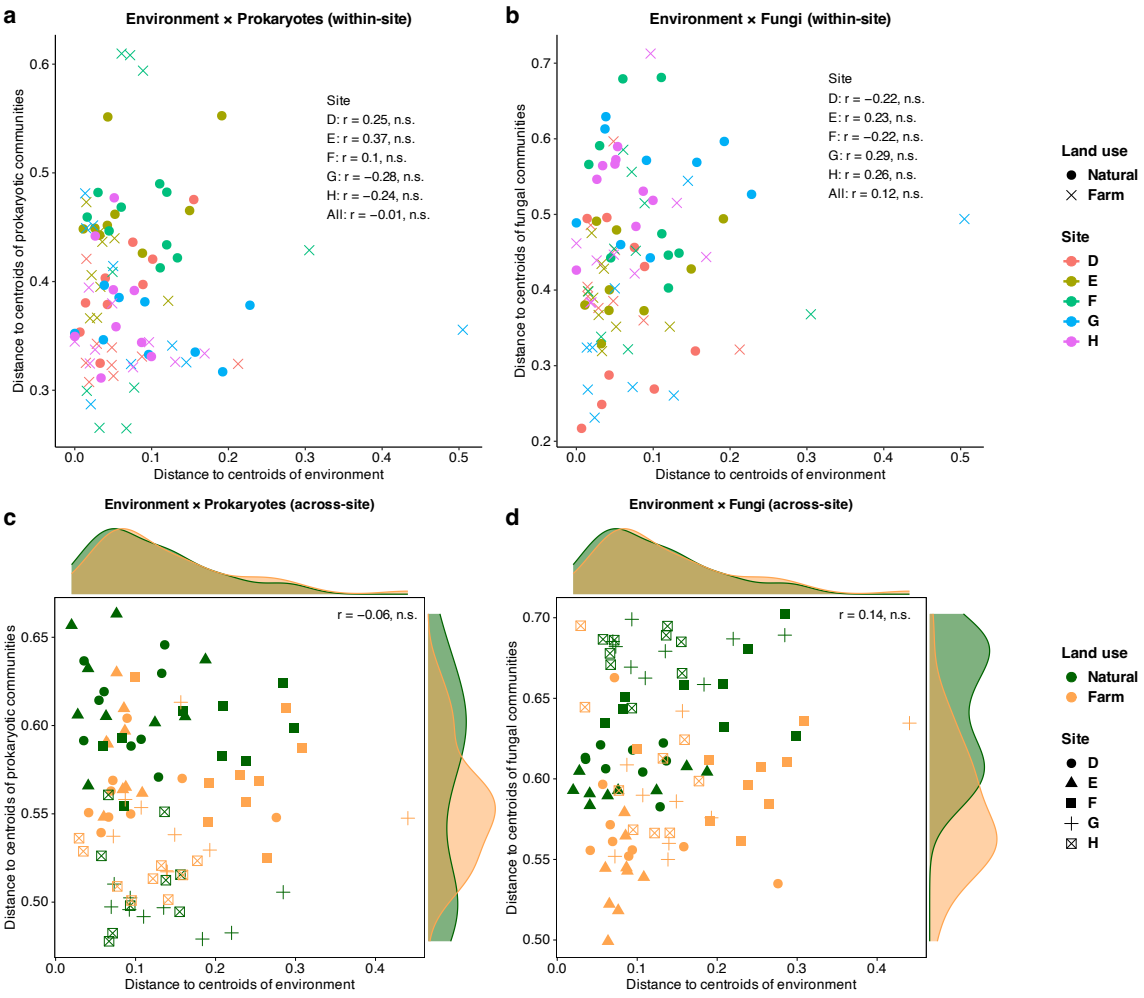

Figure S8. Relationships between heterogeneities of environmental factors and microbial communities. The correlations between distance to centroids of the environmental factors and that of microbial communities in each land use within site for (**a**) prokaryotes and (**b**) fungi, and across sites for (**c**) prokaryotes and (**d**) fungi are shown. The correlation coefficients and p-values in the Pearson's correlation tests are indicated.

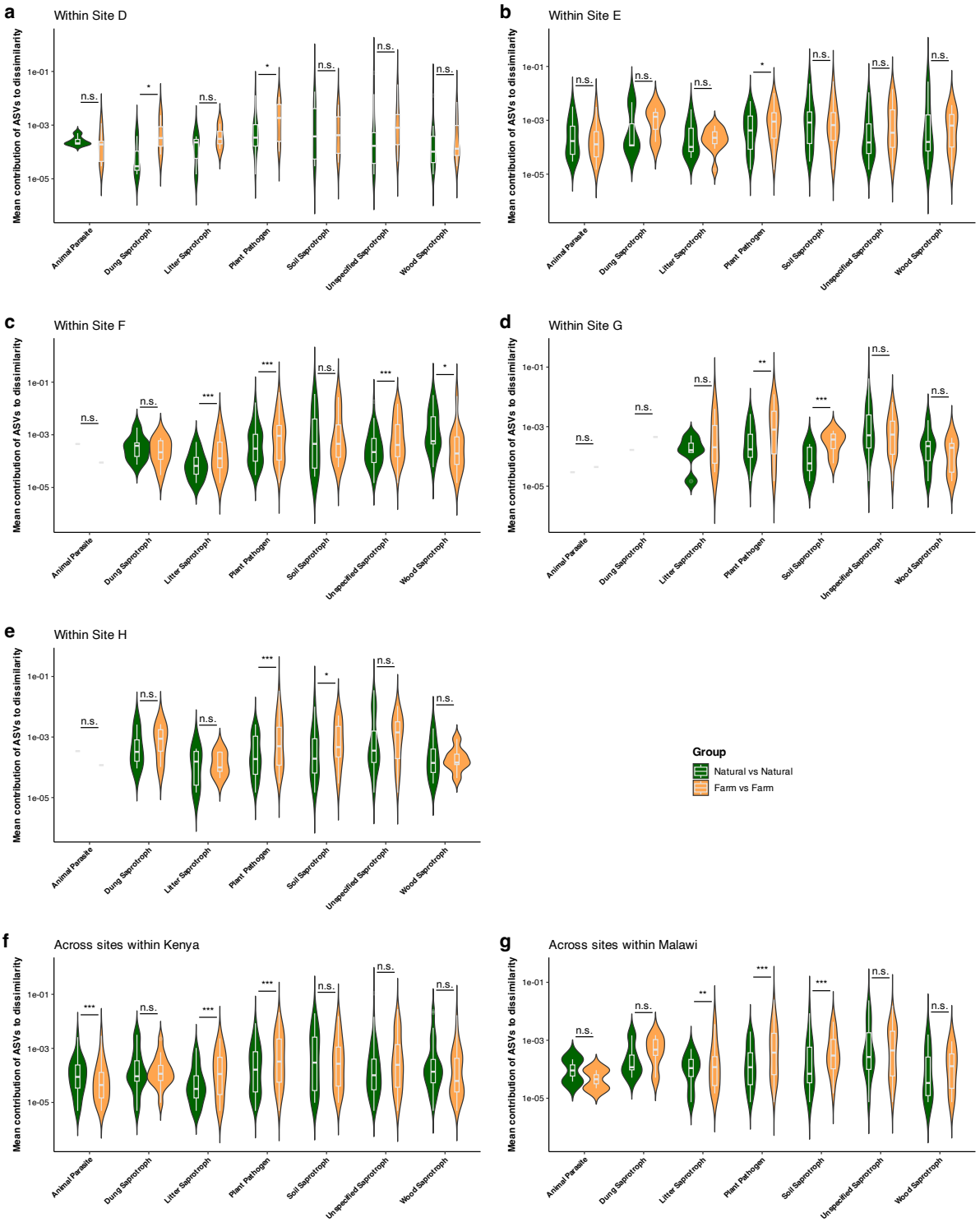

Figure S9. Contributions of fungal ASVs to community heterogeneity. The contributions of each fungal ASV to the Bray–Curtis dissimilarity among samples within each land use within site D–H (**a–e**) and within Kenya (**f**) and within Malawi (**g**) were averaged and grouped by fungal lifestyle. Asterisks ("*", "**", or "***") and n.s. indicate the significance levels of adjusted p-values of < 0.05, $p < 0.01$, $p < 0.001$, or no significant difference respectively, tested by 1000 permutations to assess differences in the mean values between natural lands and farmlands.

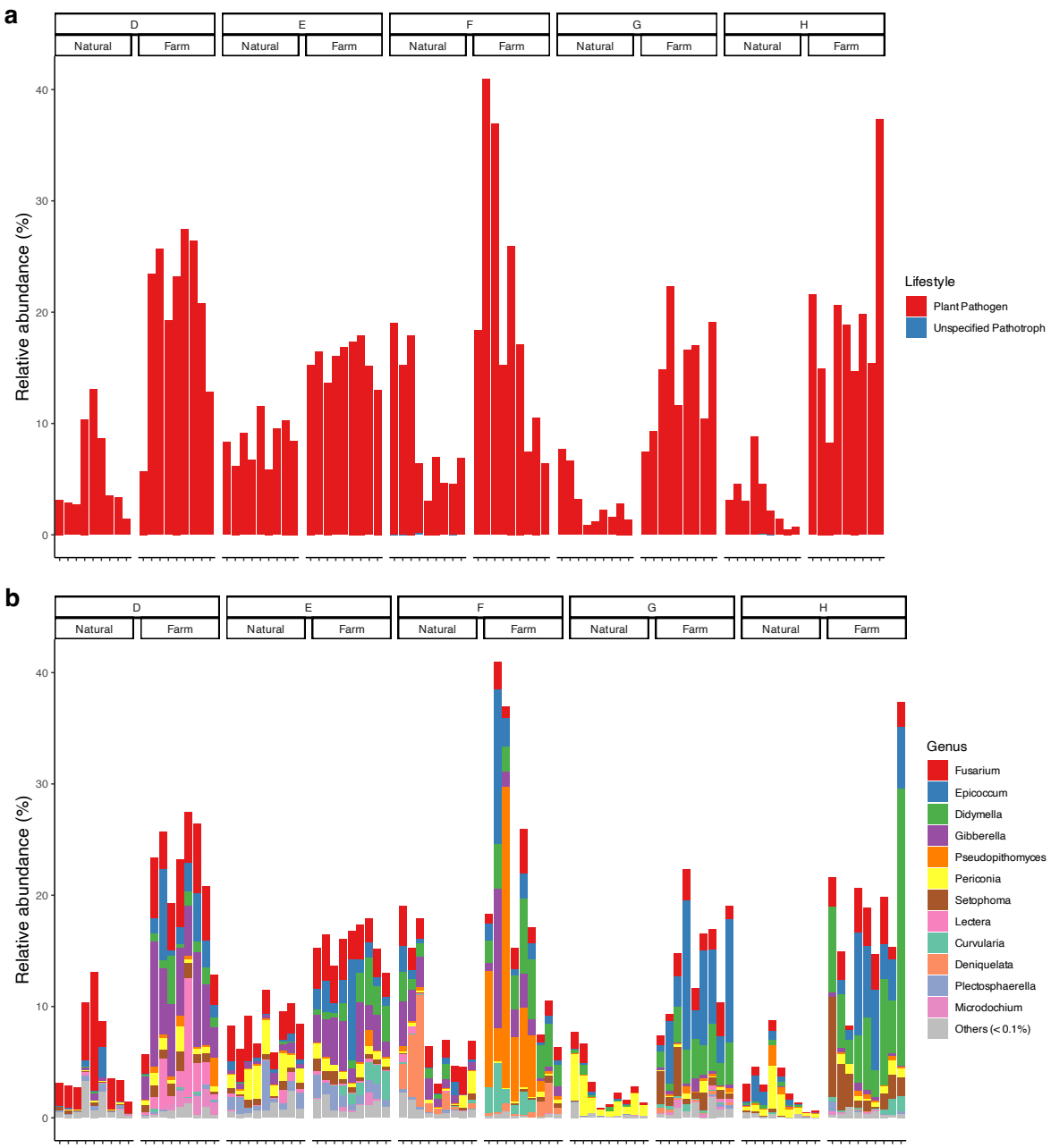

Figure S10. Composition of pathotrophs. The relative abundances of pathotrophs by primary lifestyles (**a**) and by genera that harbored mean abundances of > 0.1% (**b**) are shown.

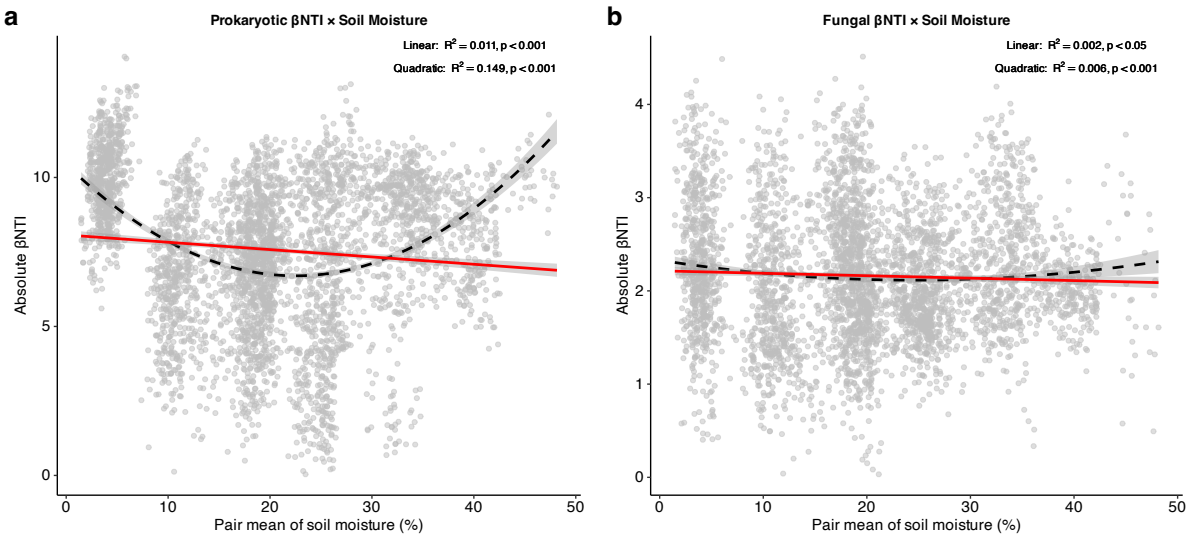

Figure S11. Relationship between soil moisture and βNTI. (**a**) Prokaryotic absolute βNTI and (**b**) fungal absolute βNTI are shown in relation to the mean soil moisture of corresponding sample pairs. The red solid line represents the linear regression model, while the black dashed line indicates the quadratic regression model. The gray shaded area corresponds to the 95% confidence interval. R^2 values and p-values for both linear and quadratic regressions are indicated.