

Supporting Information

Soil bacterial diversity is associated with human population density in urban greenspaces

Authors: Haitao Wang^{1,2}, Minying Cheng³, Melissa Dsouza^{1,4}, Pamela Weisenhorn^{1,5}, Tianling Zheng^{2,6}, Jack A. Gilbert^{1,4,5*}

¹ The Microbiome Center, Department of Surgery, University of Chicago, Chicago, IL 60637, USA

² School of Life Sciences, Xiamen University, Xiamen, Fujian Province 361102, China

³ School of Architecture, South China University of Technology, Guangzhou, Guangdong Province 510641, China

⁴ The Marine Biological Laboratory, Woods Hole, MA, 02543, USA

⁵ The Microbiome Center, Argonne National Laboratory, Lemont, IL 60439, USA

⁶ Deceased

* Corresponding Author

Phone: +1 6309152383; e-mail: gilbertjack@uchicago.edu.

Fifteen pages

Fourteen figures

Two tables

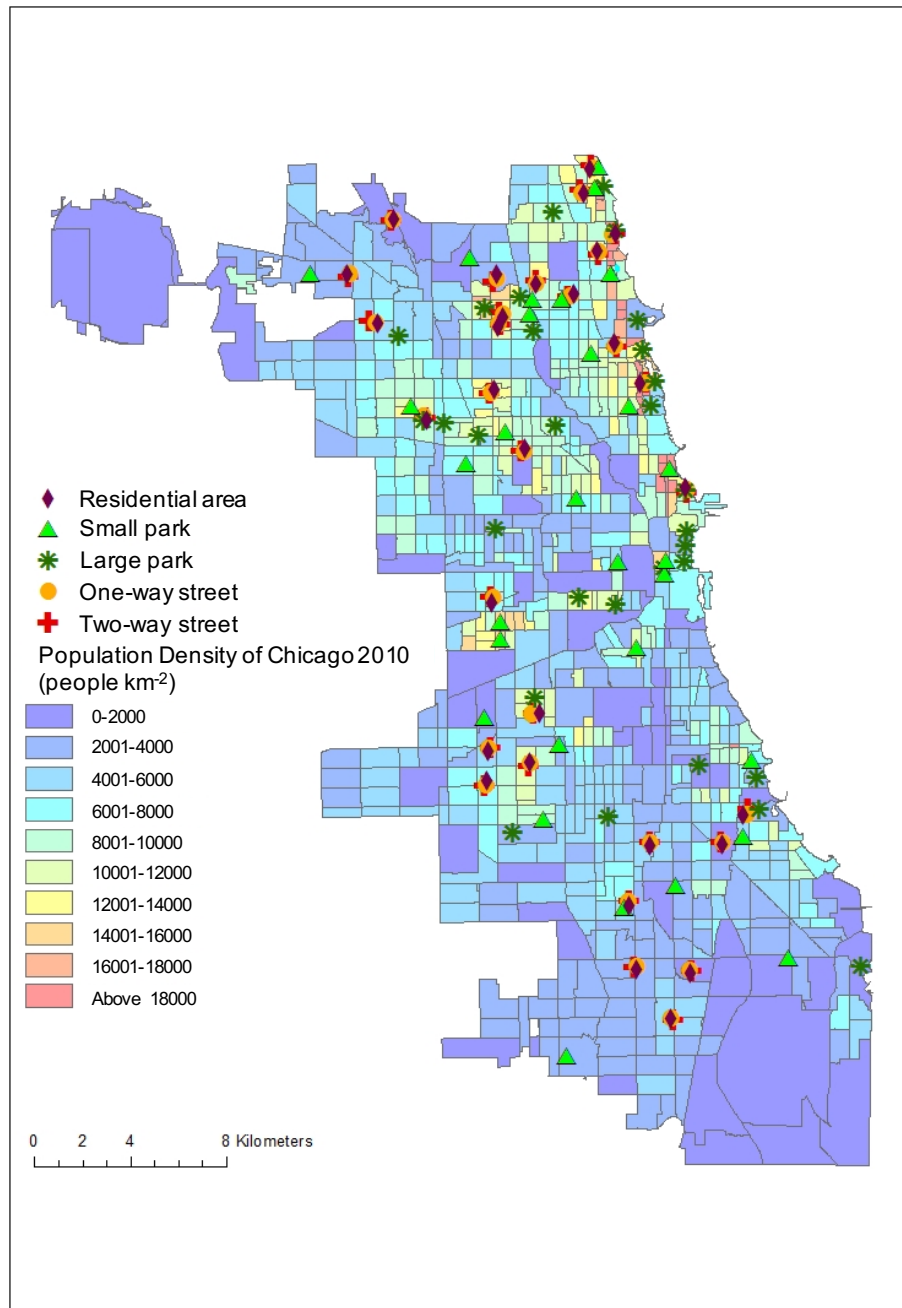


Figure S1 The human population density and distribution of sampling sites across the Chicago City.

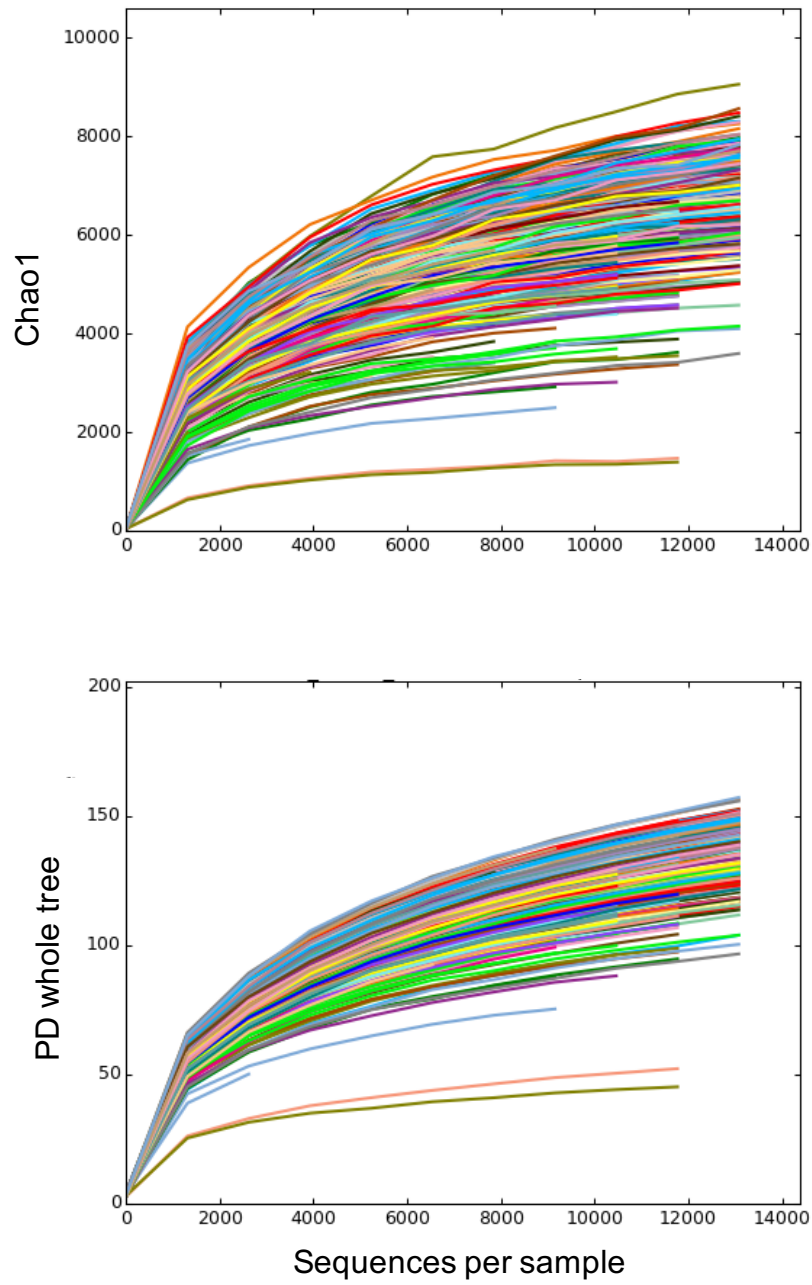


Figure S2 The rarefaction curves based on chao1 and PD whole tree (Faith's PD) indices. Each line represents one sample.

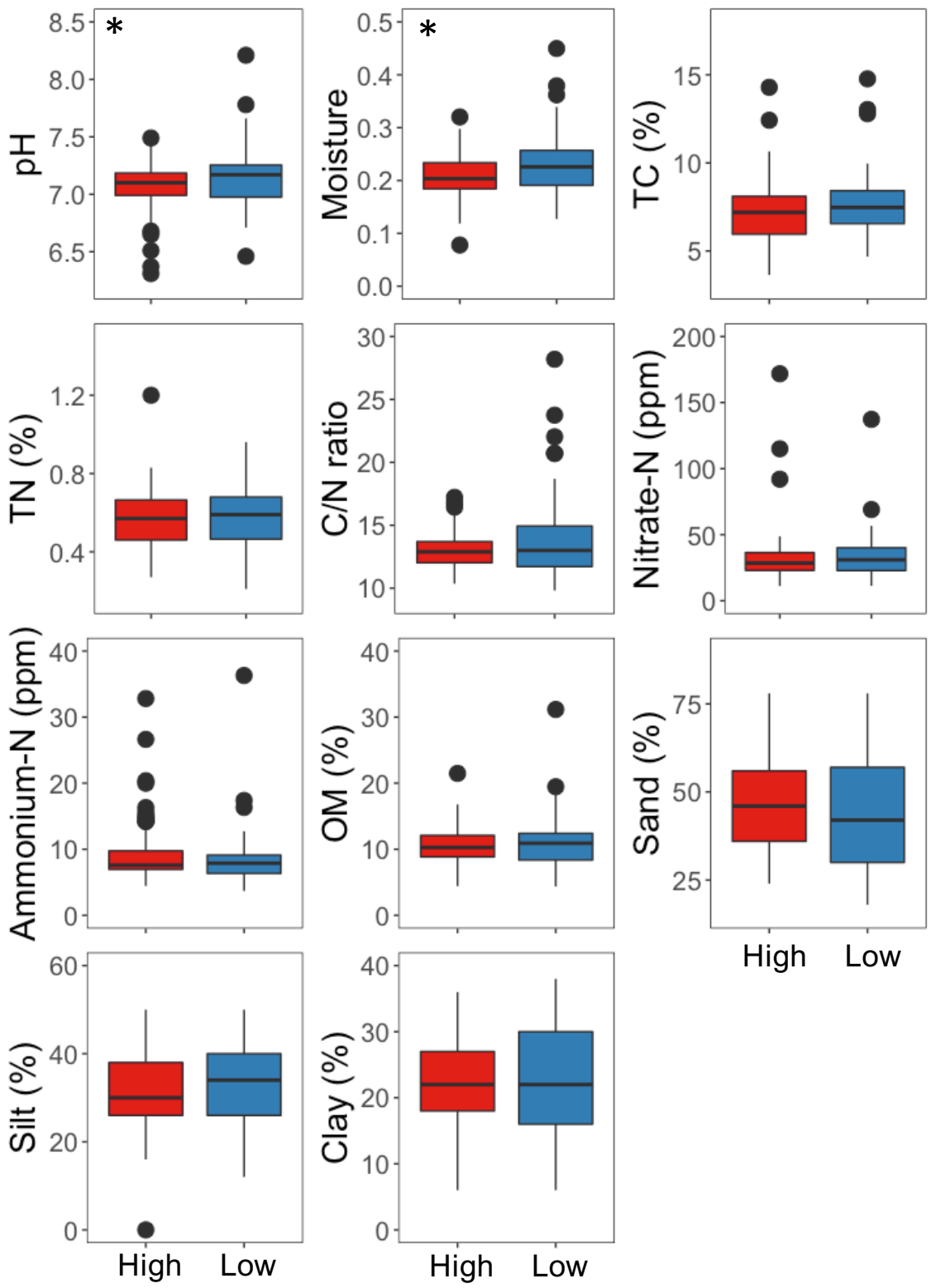


Figure S3 Boxplots showing the distribution of different soil properties in samples associated with high and low population density. *, significant differences ($P < 0.05$).

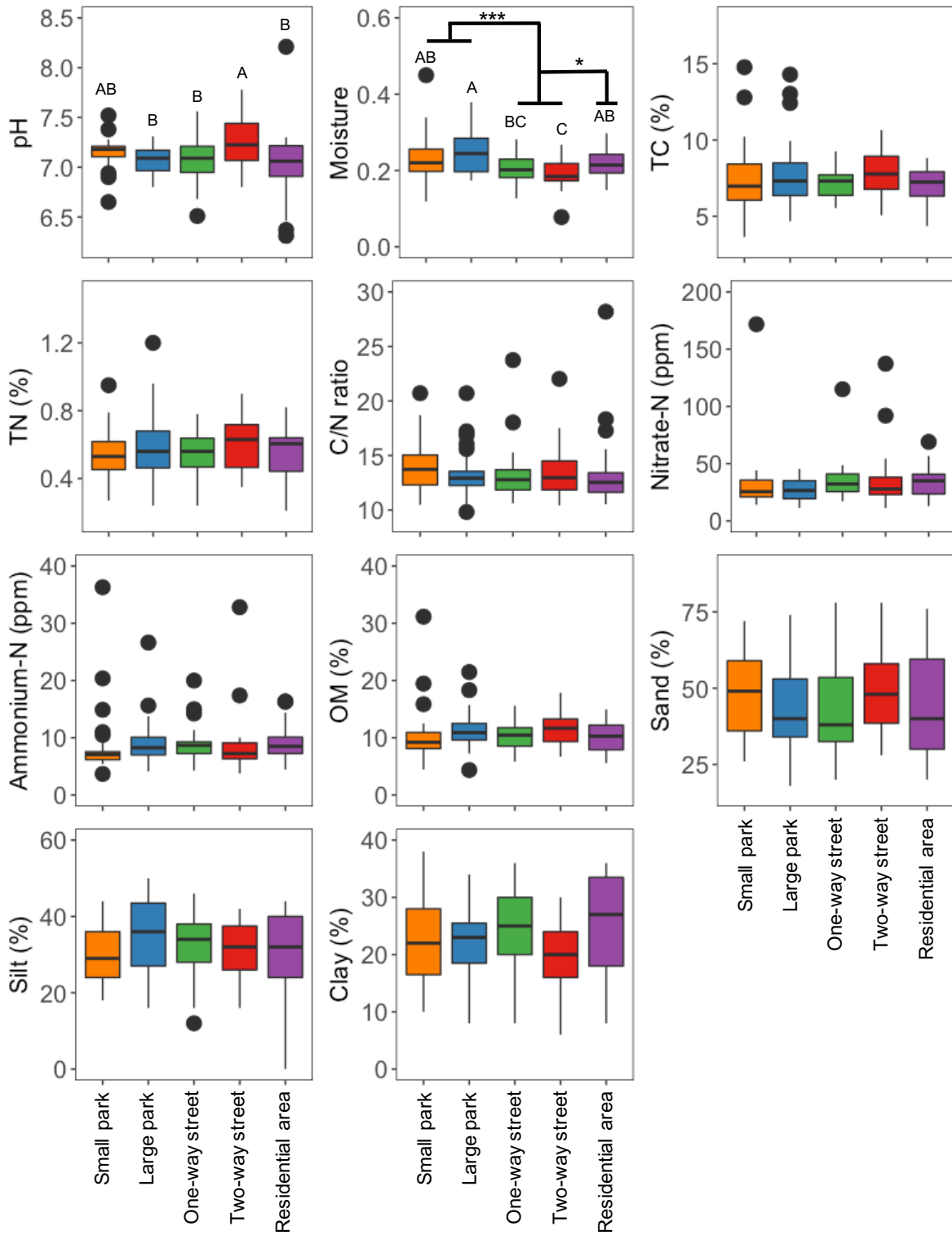


Figure S4 Boxplots showing the distribution of different soil properties in samples associated with different greenspace types. Differences are significant when no same letter exists between groups ($P < 0.05$). *, $P < 0.05$; ***, $P < 0.01$.

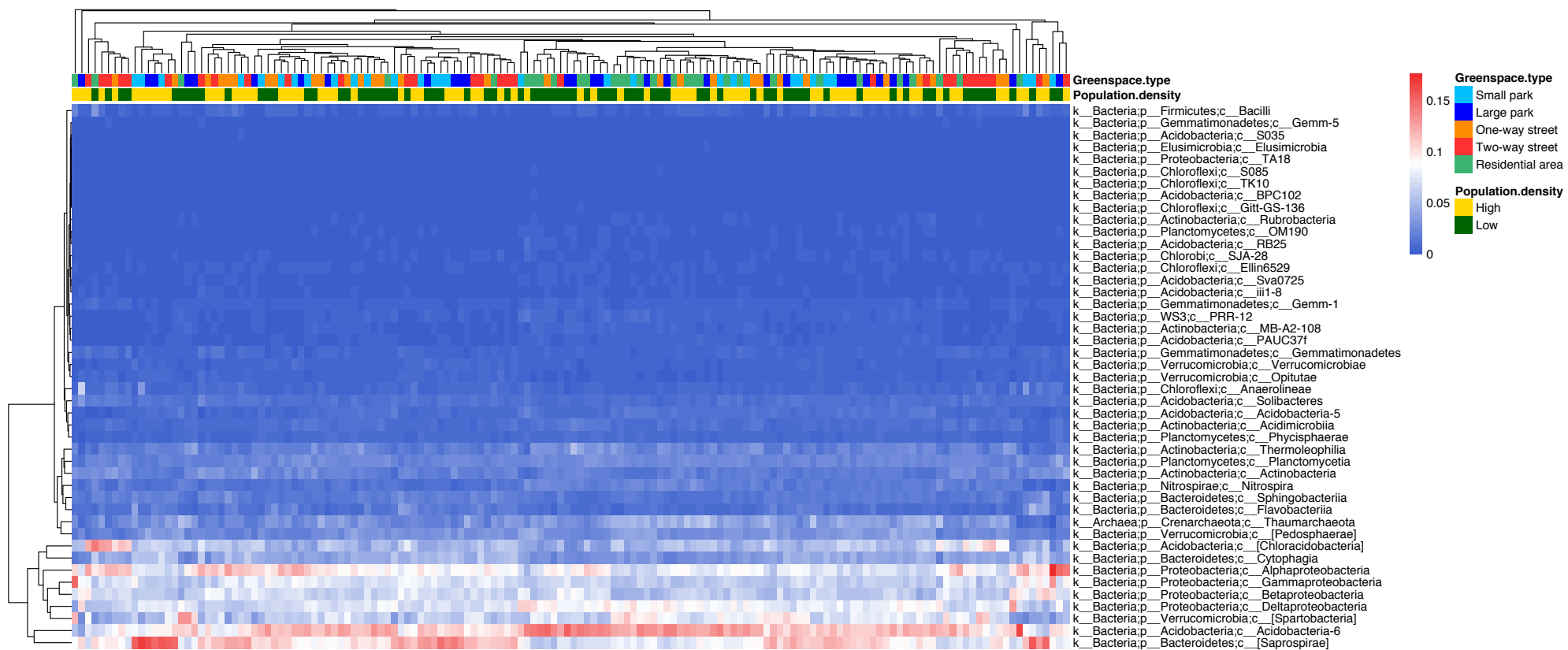


Figure S5 The distribution of abundant taxa in each site on class level. Taxa accounting for over 0.1% of the total sequences were shown.

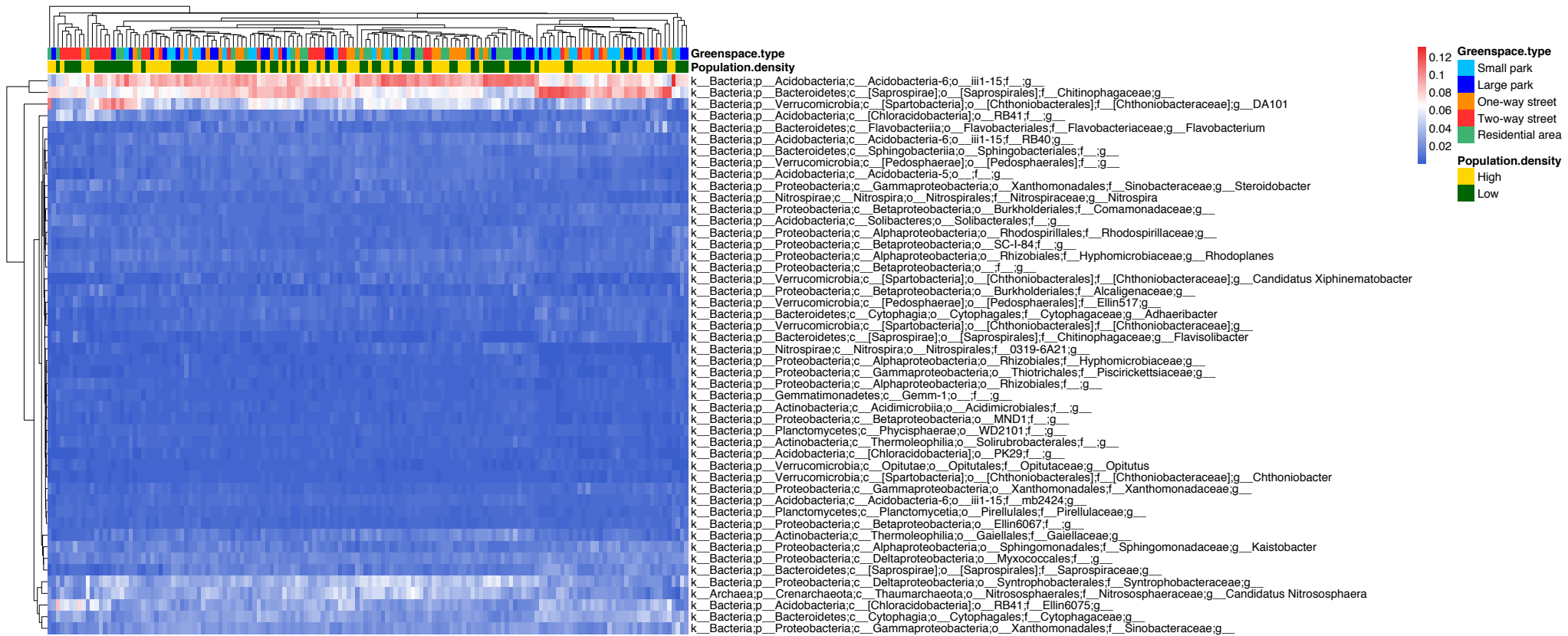


Figure S6 The distribution of abundant taxa in each site on genus level. Taxa accounting for over 0.4% of the total sequences were shown.

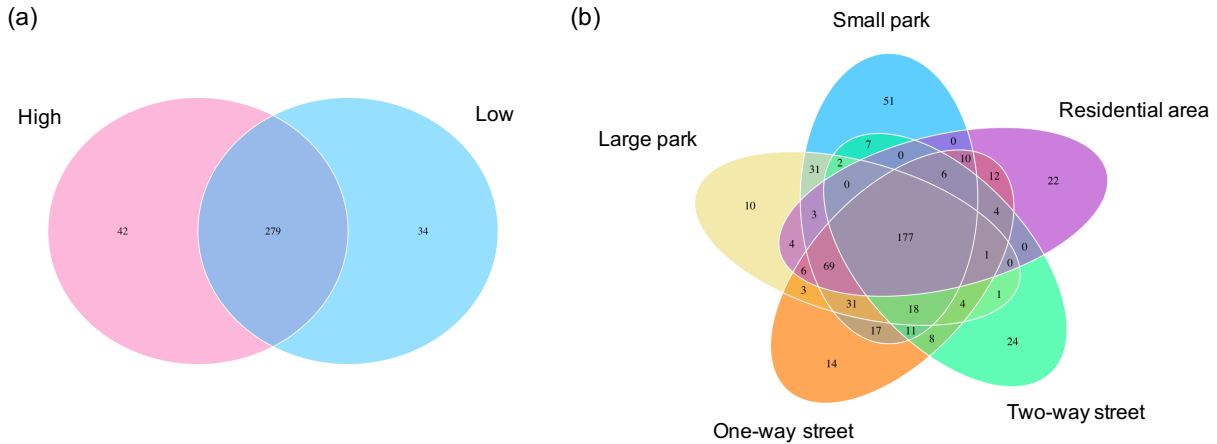


Figure S7 Venn diagrams showing the shared core microbiomes associated with different groups of human population density (a) and greenspace type (b).

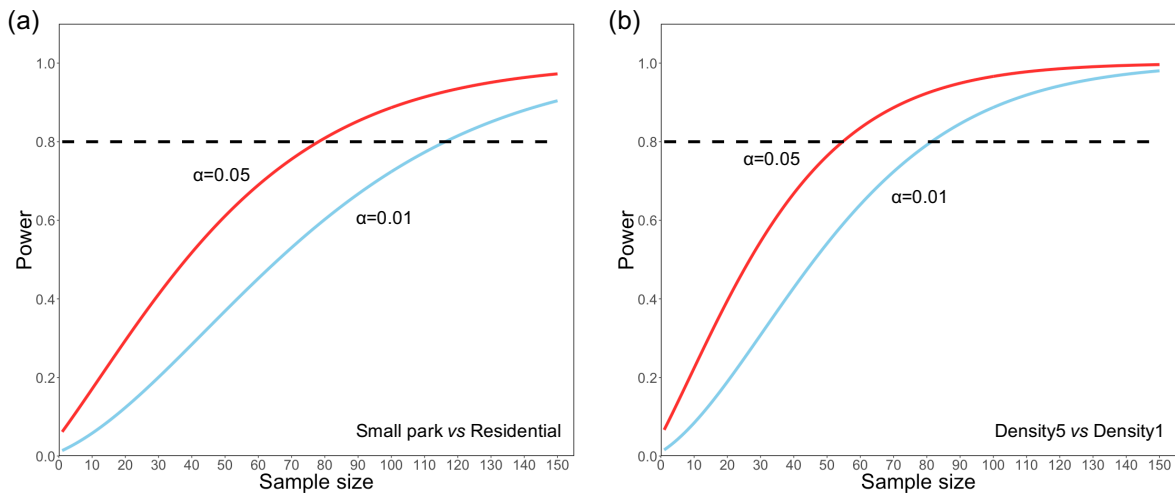


Figure S8 The statistical power levels associated with different sample sizes (per group) based on the Shannon diversity by comparing the means between small park and residential area (a) or between density5 and density1 (b) at significant levels of 0.05 (red line) and 0.01 (blue line).

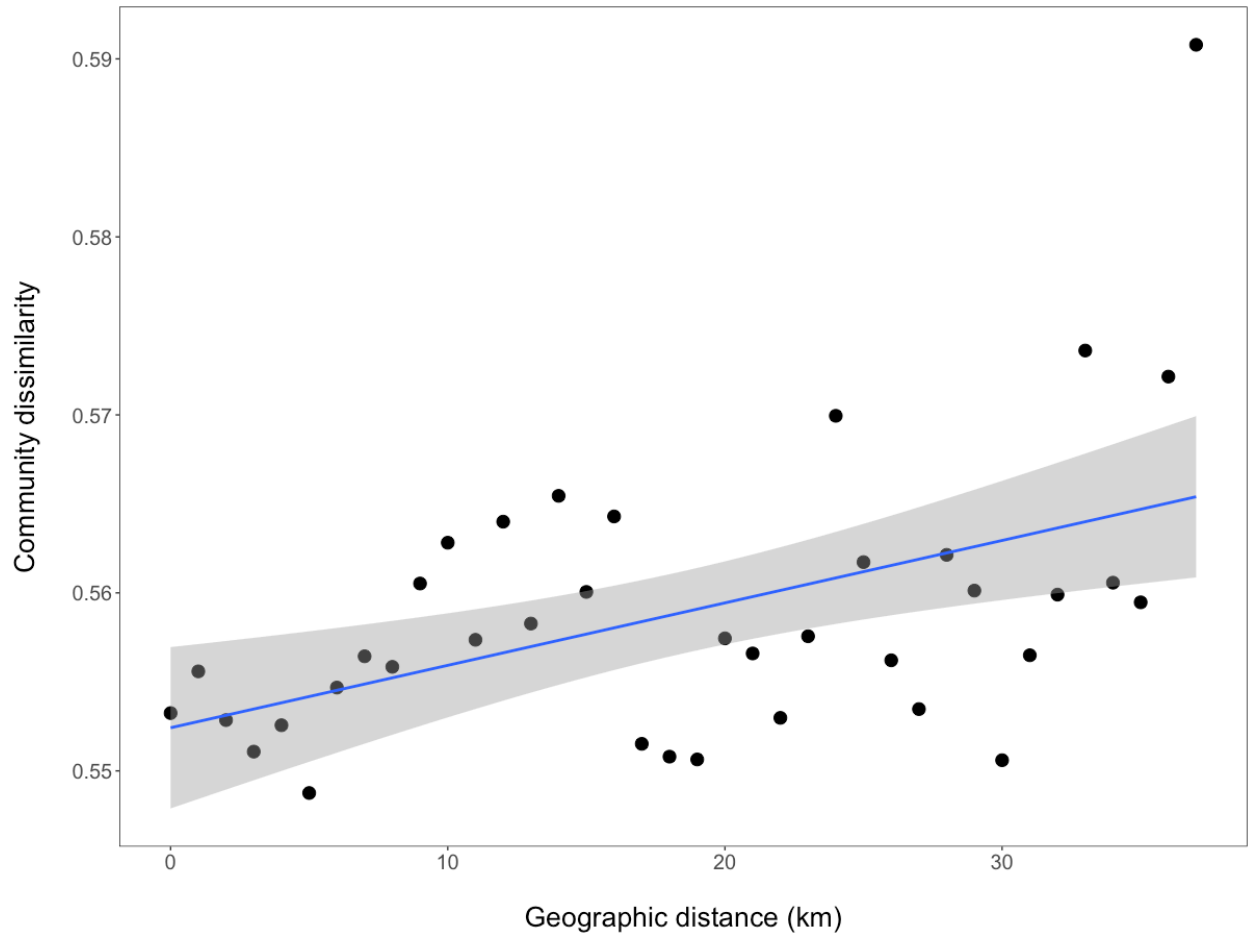


Figure S9 Relationship between the Bray-Curtis dissimilarities and geographic distances. The shade area shows the 95% confidence interval.

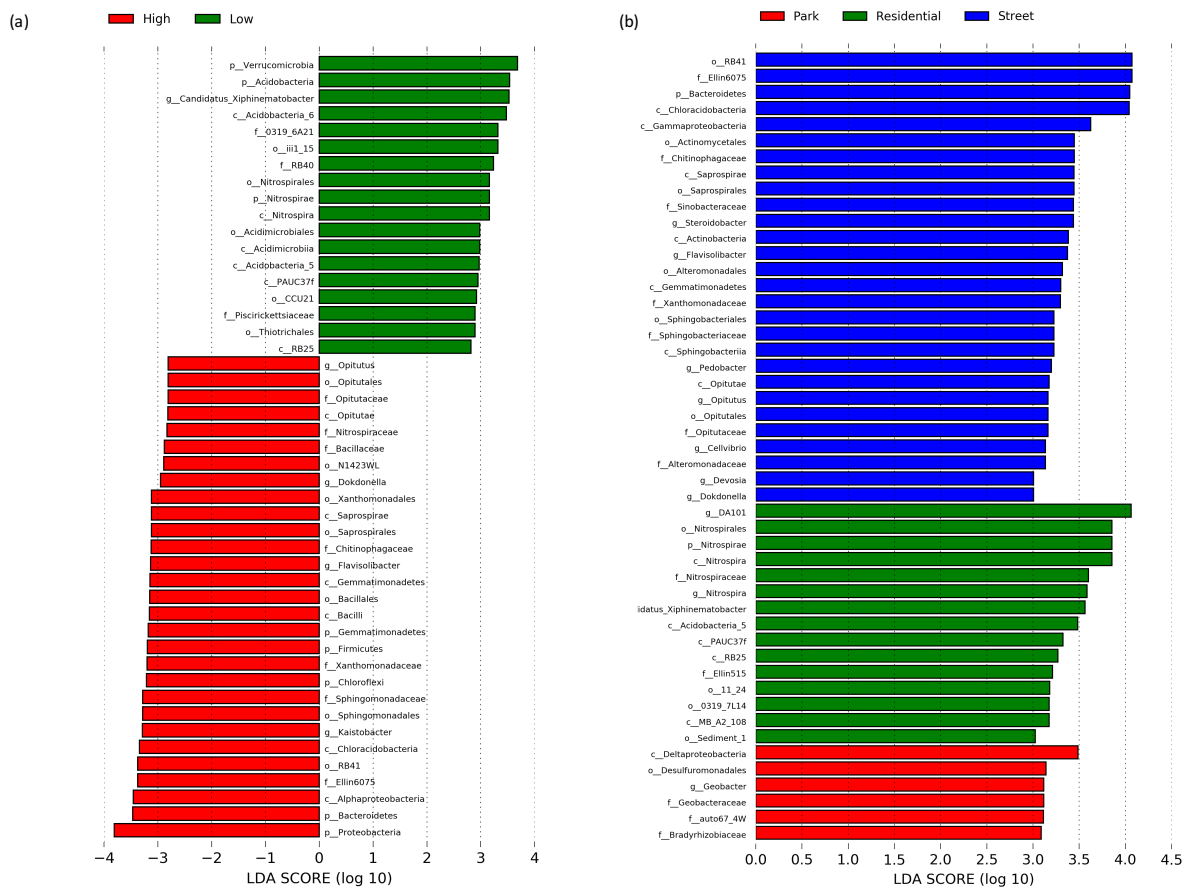


Figure S10 Histogram of the LDA scores calculated for taxa differently abundant in high and low population density soils (a) or in park, street and residential soils (b).

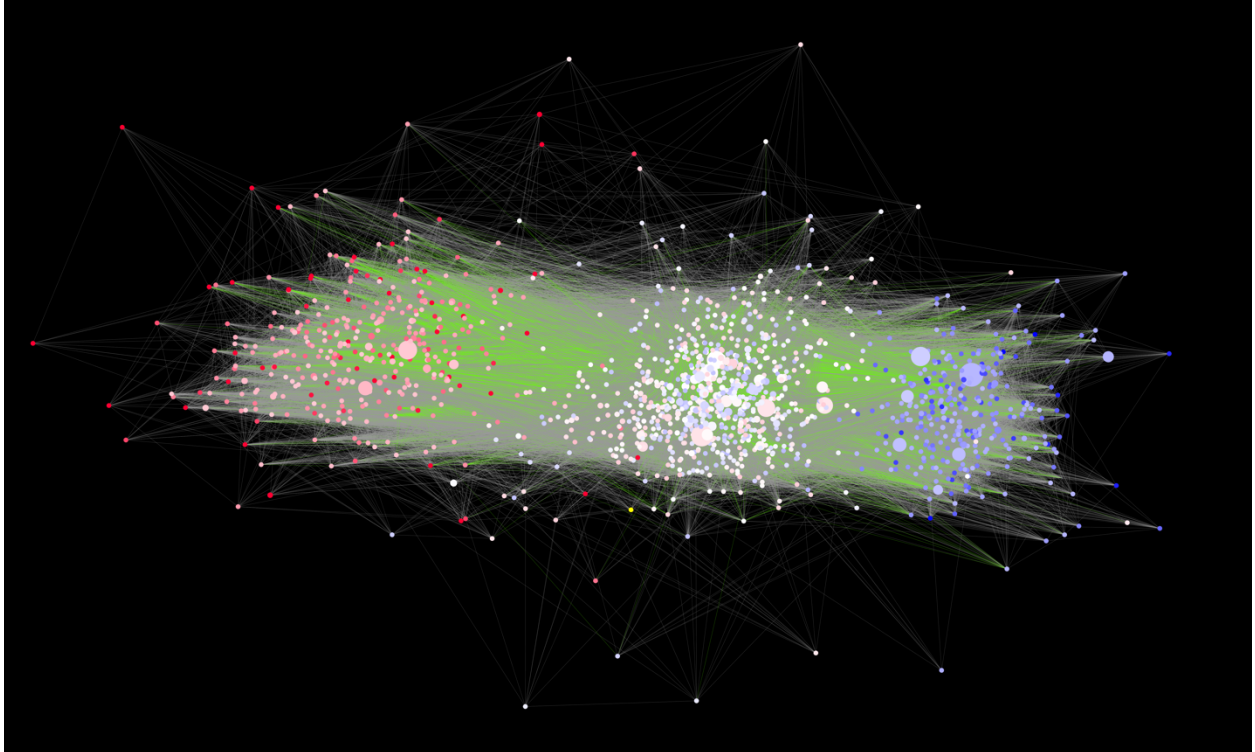


Figure S11 The co-occurrence network interactions of abundant MED nodes. The connection stands for a strong (Spearman's $\rho > 0.67$) and significant (P -value < 0.001) correlation. The vertices represent unique nodes in the data sets. The size of each vertex is proportional to the relative abundance. The scaled color represents the density ratio (the average abundance of the node in high population density samples to that in low population density samples). Red represents higher values (-1~0) while blue represents lower values (0~1). The white represents value of 1. Grey and green edges represent positive and negative correlations, respectively.

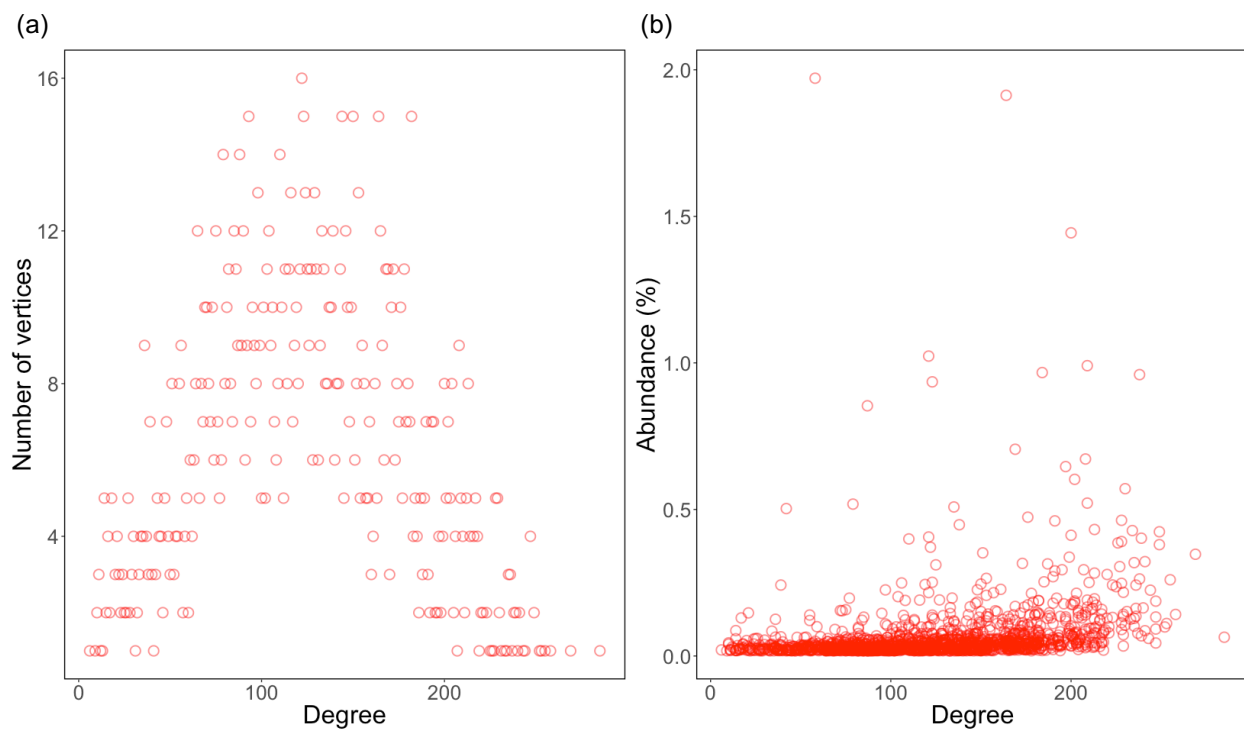


Figure S12 Scatterplots showing distribution of degrees (a) and the relationship between degree and relative abundances of the MED nodes (b).

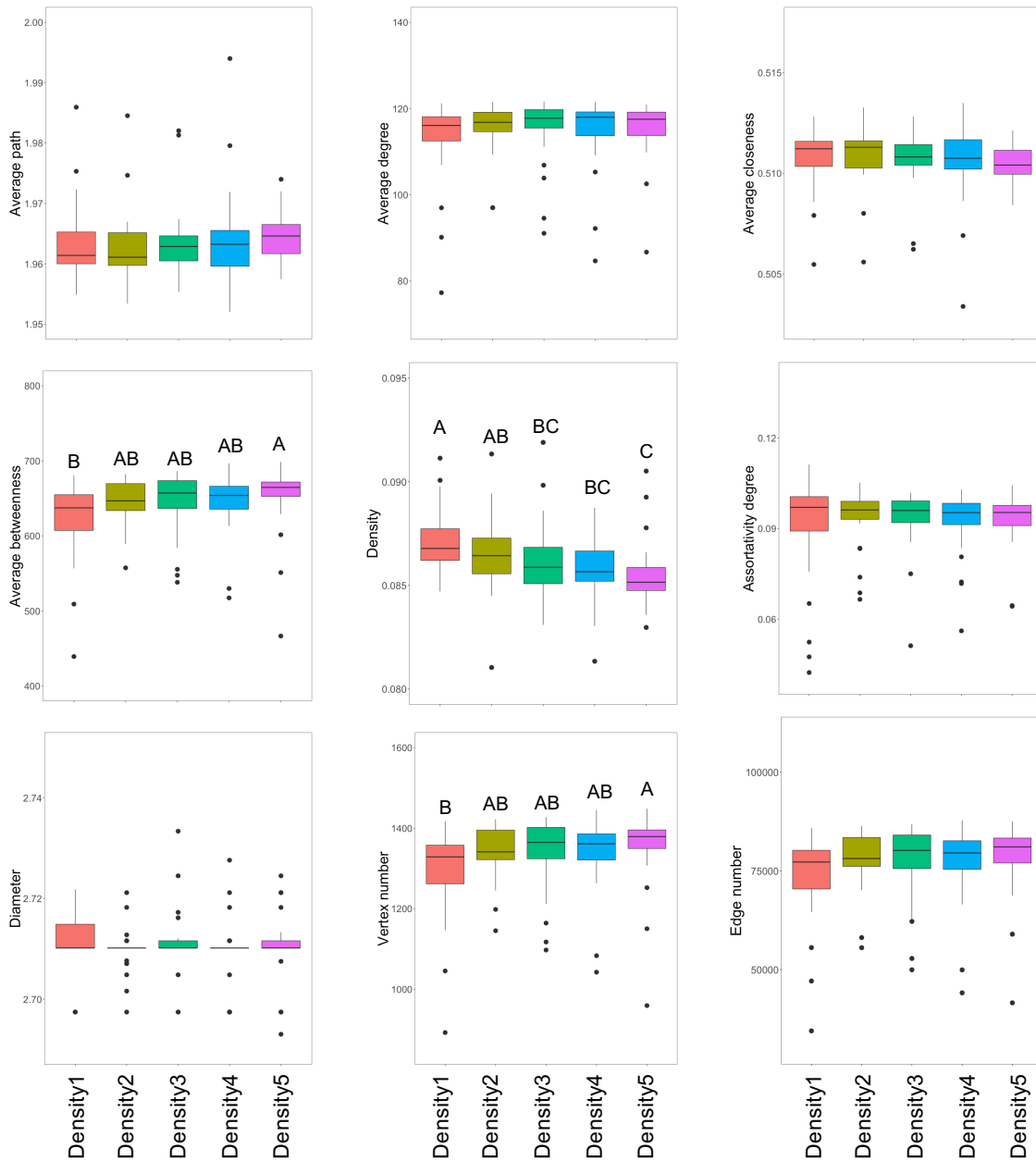


Figure S13 Boxplots showing the distribution of different network topological features in samples associated with different population density levels. Differences are significant when no same letter exists between groups ($P < 0.05$).

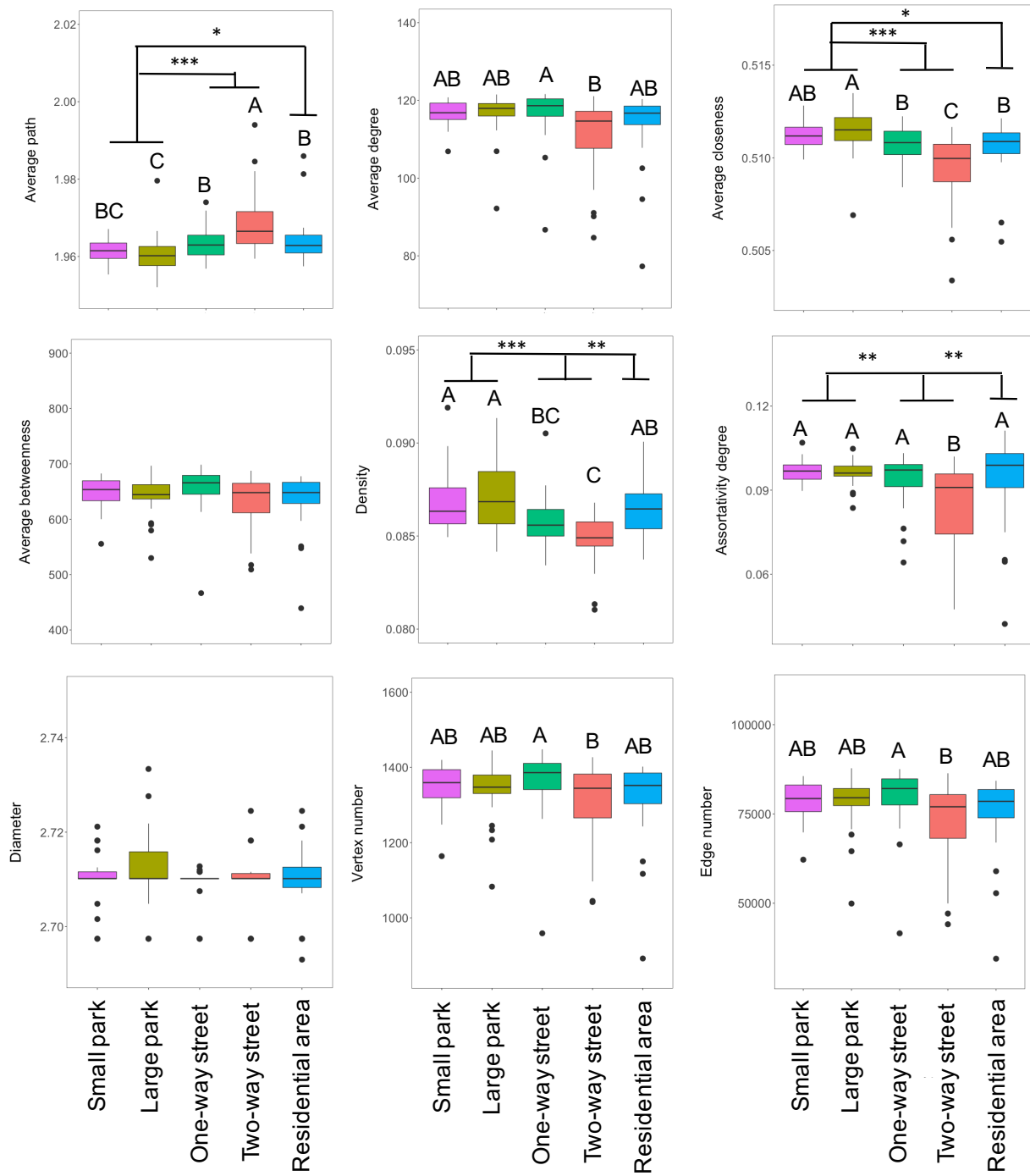


Figure S14 Boxplots showing the distribution of different network topological features in samples associated with different greenspace types. Differences are significant when no same letter exists between groups ($P < 0.05$). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Table S1 The definition of vertex-level and network-level topological features

Level	Topological feature	Definition
Vertex	Degree	The number of a vertex's adjacent edges.
	Closeness centrality	The number of steps required to access any other vertex from a given vertex.
	Betweenness centrality	The number of shortest paths going through a vertex.
	Transitivity	The probability that the neighbors of a vertex are connected.
Network	Average path	The average path length of shortest paths between all pairs of vertices.
	Average degree	The mean of the degree values for all vertices.
	Average closeness	The mean of the closeness centrality values for all vertices.
	Average betweenness	The mean of the betweenness centrality values for all vertices.
	Edge density	The ratio of the number of edges to the number of possible edges.
	Assortativity degree	The level of homophily of the graph based on some vertex labeling or values assigned to vertices.
	Diameter	Length of the longest geodesic.
	Vertex number	The number of vertex or MED nodes in a network.
	Edge number	The number of connected edges in a network.

Table S2 The importance of soil properties, human population density and greenspace type for microbial community structure using MRM model

Variables	Coefficient	<i>P</i> value	Significance level
pH	0.212	0.001	***
Moisture	0.081	0.001	***
TC	0.089	0.001	***
TN	-0.102	0.001	***
C/N ratio	0.034	0.085	NS
Nitrate	-0.005	0.765	NS
Ammonium	0.082	0.001	***
OM	0.090	0.001	***
Sand	0.077	0.008	**
Clay	0.028	0.163	NS
Silt	-0.045	0.102	NS
Population density	0.062	0.001	***
Greenspace type	0.079	0.001	***

** , $P < 0.01$; *** , $P < 0.001$. NS, not significant.