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

Haitao Wang[1,2], Minying Cheng[3], 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.

Note

The authors declare no competing financial interest.

Abstract

Urban greenspaces provide extensive ecosystem services, including pollutant remediation, water management, carbon maintenance, and nutrient cycling. However, while the urban soil microbiota underpin these services, we still have limited understanding of the factors that influence their distribution. We characterized soil bacterial communities from turf-grasses associated with urban parks, streets and residential sites across a major urban environment, including a gradient of human population density. Bacterial diversity was significantly positively correlated with the population density; and species diversity was greater in park and street soils, compared to residential soils. Population density and greenspace type also led to significant differences in the microbial community composition that was also significantly correlated with soil pH, moisture and texture. Co-occurrence network analysis revealed that microbial guilds in urban soils were well correlated. Abundant soil microbes in high density population areas had fewer interactions, while abundant bacteria in high moisture soils had more interactions. These results indicate the significant influence of changes in urban demographics and land-use on soil microbial communities. As urbanization is rapidly growing across the planet, it is important to improve our understanding of the consequences of urban zoning on the soil microbiota.

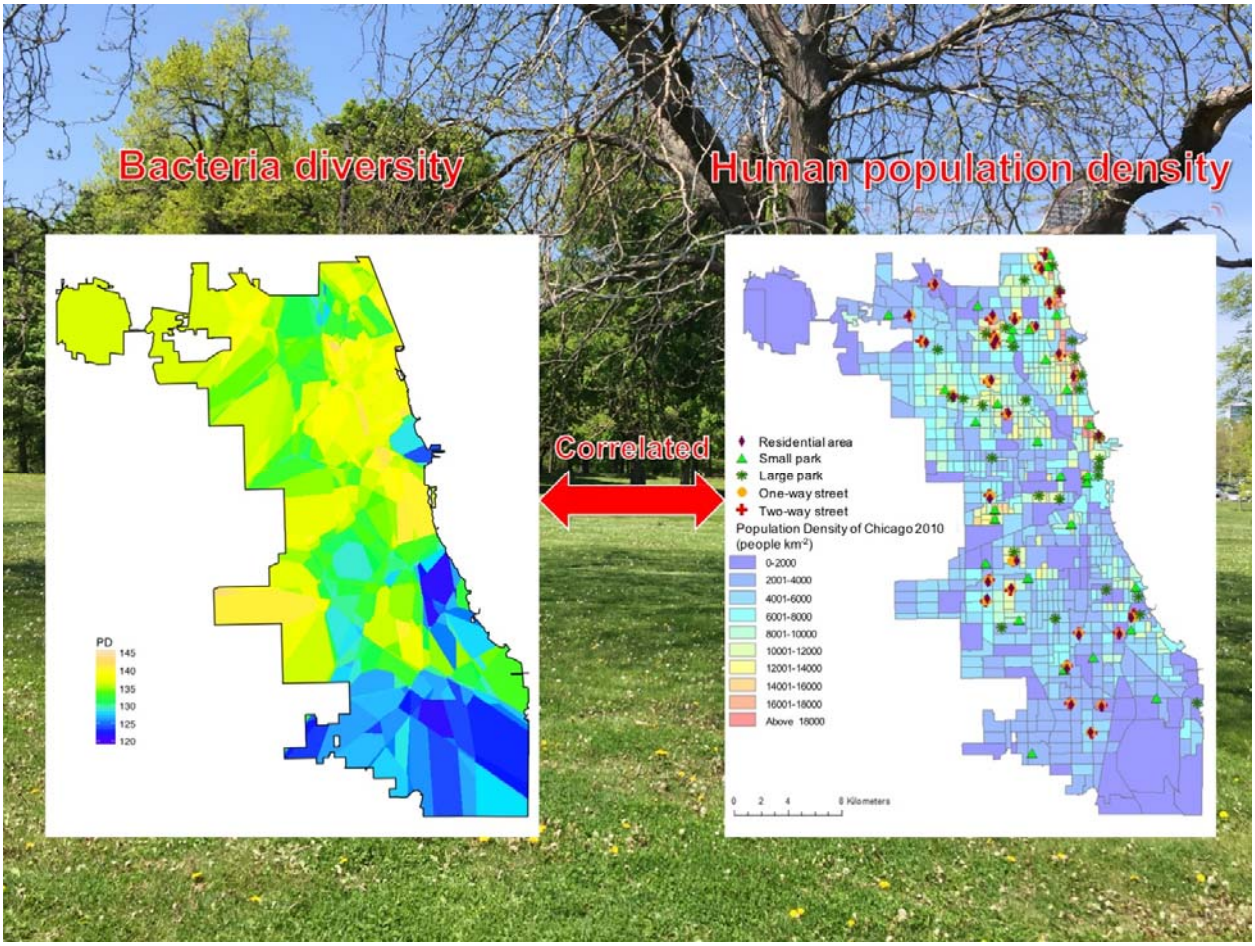


Table of Contents (TOC)/Abstract Art

Introduction

Urban spaces are expanding at an unprecedented rate, and this transition of human populations from rural to urban living increases material production demands and human consumption, which can have dramatic effects on land use, leading to dramatic shifts in ecosystem function.[1,2] Understanding how these disturbances influence soil microbial communities lays the foundation for maintaining soil quality and health, and for understanding the influence of urbanization on both global biogeochemical cycles and local nutrient cycles.[3-6] Urban greenspaces are a vital component of the urban ecosystem performing essential ecosystem services, including pollutant remediation,

water management, carbon maintenance, and nutrient cycling.[7-10] And the urban environment alters the microbiota of soils, which leads to changes in soil microbial functions. Urban greenspace soils can also be a major reservoir of genetic and species diversity,[11,12] and it is therefore important to catalogue this resource and understand the impact of soil physicochemistry, and major urban factors on soil microbial communities in urban greenspaces.

Land-use type could be an important factor that shapes the microbial distribution and functions in urban area. Urbanization contributes to frequent land management in built areas that differentiates land-use types for different purposes. For example, turf-grasses, which dominate the urban greenspace in North American cities,[13] comprise the lawns of residential, commercial and institutional areas, sports fields, and golf courses.[14] Turf-grass is also employed by city authorities for street medians and parks and is subjected to regional management strategies that allow it to be productive, and therefore independent of the local climate.[8] Since different greenspaces demand different management strategies, understanding the microbial variance across different urban turf-grass sites could lead to precision management, improving soil quality and health.

Human population density, as a proxy of anthropogenic activity, could also impact the underground communities. As a typical symbol of urbanization, the increase in human population density always companies with the increase in species richness.[15] For example, some taxonomic groups were found to maintain a positive relationship between population density and species richness.[16-19] However, some studies suggest that the growth of human population and population density poses threats to animal and plant biodiversity.[20,21] Despite of the ambiguous results, it is worth noting that while these findings mostly focus on the plants and animals, the most diverse community, microbes, is often not studied. The relationship between population density and microbial biodiversity at a whole city scale remains unknown.

Understanding the influence of these factors on microbial distribution and functions across a complex urban environment, requires detailed characterization of human activity, land use, physicochemistry and population density. As the third largest city of United States, Chicago has a long legacy of both industrial activity and investment in public parks.[22] The parks range from extensive contiguous parkland to a network of city-managed neighborhood parks surrounding by residential zoning. The greenspaces in Chicago offer an opportunity to determine the influence of park, street and residential zoning, and human population density in a well-established urban ecosystem. We performed 16S rRNA amplicon sequencing at sites across this urban area over a gradient of land use and human population density. Recent studies suggest that co-occurrence patterns of soil microbiota could provide novel insights into potential microbial interactions and illustrate niche spaces within certain community members.[23] However, the soil microbial co-occurrence network across an urban area remains under-characterized. We therefore characterized the geographic patterns of microbial diversity, community composition and topologic features of microbial co-occurrence network. We aimed to investigate the associations between these patterns and human population density or land-use type. Specifically, we aimed to reveal the spatial distribution of microbial diversity and its relationship with human population density at a whole city scale. This study intends to provide a theoretical guide for future urban planning and land management.

Materials and methods

Study sites and soil sampling

Soil samples were collected between 23 May and 30 May 2016 from five different greenspaces in the City of Chicago. The weather was mostly sunny and there was no rainfall during the sampling

dates. The temperature ranged from 21 ºC to 29 ºC during the daytime. Sampling sites were chosen according to the population density of the city as shown in Supporting Information (SI) Figure S1. The population density was calculated by the human population to the area size of each census which were accessed from the City of Chicago's Official Site (https://www.cityofchicago.org/city/en.html). Five greenspace types (small park, large park, one-way street, two-way street and residential area) were selected and samples were taken across the 10 population densities (SI Figure S1). For each density, three censuses were chosen as replicates (SI Figure S1). In total, 30 censuses associated with 10 population densities were chosen, and in each census, 5 sites representing five greenspace types were chosen (SI Figure S1). In total, there were 150 sampling sites. Parks were divided into small and large based on their sizes and functionalities. The small parks were normally located in the residential area with a size lower than 0.1 km^2. Most of the small parks we selected were named as playground or playlot park land according to the Chicago Park District. The large parks had more functionalities including playground, sports field, watershed and concession stands. For example, Grant Park, with a size of 1.29 km^2, is in the center of the city and serves as a performance venue, gardens, art work, sporting, and harbor facilities. Therefore, there were differences in anthropologic influences on the soil communities between small and large parks. All the park information was from the website of Chicago Park District (http://www.chicagoparkdistrict.com/). The street soils were from the street green belts. The car flow was higher in the two-way street compared to the one-way street, but absolute quantification of car flow, or exhaust pollution was not monitored. The residential soils were associated with private lawns of the residents who lived close to a one-way street. All the street information was from the Google Maps (https://www.google.com/maps/). These five

119 greenspace types are subject to varied urban stresses, as a result of differing land usage and
120 management, and can be used as proxies of land-use types.
121 At each site, we collected 5 replicate samples at a distance of 10 m from each other in street
122 green belts and residential lawns while samples were taken in a cycle in parks at the same distance.
123 Sampled sites avoided trees and obvious tree root systems. A 30-mL sterilized syringe, with tip
124 removed, was inserted in the surface soil to a depth of 10 cm, which was then removed and the
125 soil ejected into a sterilized Nasco Whirl-PAK baggie. Then the soils were immediately blended
126 and transported to the laboratory using an ice box. After grass roots were removed with a sterilized
127 tweezer, soils used for DNA extraction were frozen at -80 ºC on the same day of collection. In
128 total, 750 soil samples (10 population densities × 3 population density replicates × 5 greenspace
129 types × 5 sample replicates) were collected.
130 **Soil physicochemical and edaphic properties**
131 The soil replicate samples from each sampling site were equally mixed and the soil properties were
132 then measured. Soil moisture was measured immediately after sampling each day. The moisture
133 was measured gravimetrically by drying the soil over 12 h at 105 ºC. After that, we air dried the
134 soils before sending them to the Soil Testing Laboratory in Kansas State University
135 (http://www.agronomy.k-state.edu/services/soiltesting/) to measure the other edaphic properties
136 including pH, total carbon (TC), total nitrogen (TN), nitrate, ammonium, organic matter (OM) and
137 texture. The detailed methods can be found on the website of the Soil Testing Laboratory.
138 **DNA extraction and sequencing**
139 The amplicon sequencing was conducted according to the protocol of Earth Microbiome Project
140 (http://press.igsb.anl.gov/earthmicrobiome/protocols-and-standards/16s/) with some
141 modifications. The total DNA of each sample was extracted from 0.25 g soil using the Power Soil

Kit (MoBio, USA). The V4 region of bacterial 16S rRNA gene was amplified with 515f (5'-GTGYCAGCMGCCGCGGTAA-3') and 806r (5'-GGACTACNVGGGTWTCTAAT-3') primer pair.[24] The reverse primer was tagged with a 12-base barcode sequence. Each 25 μL PCR reaction contained 12.5 μL of 2X AccuStart II PCR ToughMix (Quantabio, USA), 9.5 μL of PCR water, 1 μL of each primer (5 μM), 1 μL of DNA templete. The amplification protocol was as follows: 94 ^{o}C for 3 min; 35 cycles of 94 ^{o}C for 45 s, 50 ^{o}C for 60 s and 72 ^{o}C for 90 s; 72 ^{o}C for 10 min. The concentrations of PCR products were fluorometrically quantified using PicoGreen kit (Thermo Fisher, USA). The barcoded PCR products were then equally pooled and this pool was then purified using UltraClean PCR Clean-Up Kit (Mo Bio). The final pool was sequenced on an Illumina MiSeq platform.

Sequence data analysis

The raw paired sequence reads were joined and de-multiplexed using QIIME.[25] Sequences that failed to meet the default quality scores were discarded. Filtered sequences were de-noised using Minimum Entropy Decomposition (MED).[26] MED employs Shannon entropy to identify nucleotide variations in a sequence alignment that are used to discriminate MED nodes which represent operational taxonomic units. MED is able to produce finer scale resolution descriptions of microbial communities compared to pairwise similarity approaches. The decomposition was run with default settings with the exception of the minimum substantive abundance which was set to 15. The chimeras were checked and filtered using the UCHIME algorithm in USEARCH.[27] The sequence data clustered into 37,927 MED nodes. Then the representative sequence of each node was assigned to taxonomy against the Greengenes database.[28] Based on rarefaction (SI Figure S2), the MED node matrix was rarefied to 8,000 sequences per sample. Samples with < 8,000 sequences were discarded, resulting in a total of 719 samples that were further processed for analysis.

Alpha diversity indices including Shannon index and phylogenetic diversity (Faith's Phylogenetic Diversity) were calculated before rarefying. Core microbiota were defined as MED taxonomic units that were present in at least 80% of the samples in each group of population density or greenspace type. Beta diversity was calculated using Bray-Curtis dissimilarity scores, and visualized using non-metric multidimensional scaling (nMDS). LDA Effect Size (LEfSe) [29] was determined to identify the significant and abundant taxa associated with different groups of factors (population density and greenspace type). LEfSe analysis was carried out according to the instructions on the website (http://huttenhower.sph.harvard.edu/galaxy).

Co-occurrence network was constructed according to Ma *et al*.[23] with some modifications using R v3.3.2. MED nodes with relative abundances lower than 0.01% were removed. The pairwise Pearson correlations were conducted with the *WGCNA* package and all *P*-values were adjusted by the Benjamini and Hochberg false discovery rate (FDR) method using *multtest* package. The cutoff of adjusted *P*-value was 0.001. Deconvolution method was used to discern the direct correlation dependencies and the random matrix theory-based method was used to determine the cutoff of the modified coefficients. The coefficient cutoff was 0.67. The network image (SI Figure S11) was generated with Cytoscape 3.3.0[30] and vertex-level topological features, including degree, transitivity, closeness and betweenness centrality (SI Table S1), were assessed via *igraph* package.

The sequence data were deposited to the European Nucleotide Archive of EMBL (European Molecular Biology Laboratory). The study accession number is PRJEB21935.

Statistical analysis

All the statistical analyses were done with R. According to the distribution of population density, soil samples were evenly categorized into either two groups: low density (<10,000 people km^{-2}) and high density (>10,000 people km^{-2}) or five groups: density1 (<3,350 people km^{-2}), density2

(3,350 to 6,600 people km^{-2}), density3 (6,600 to 10,000 people km^{-2}), density4 (10,000 to 13,500 people km^{-2}) and density5 (>13,500 people km^{-2}). The Spearman's rank correlation was employed to examine the relationships between soil properties, factors (population density and greenspace type) and diversity indices as well as network topological features. The spatial distributions of the alpha diversity indices were predicted using automatic Krige interpolation in the *automap* package. The statistical power levels based on Shannon diversity were calculated to detect the robustness of sample sizes according to Chow *et al*.[31] The shared core microbiome among different groups were explored via Venn diagram by using the *VennDiagram* package. The environmental factors were fitted into nMDS ordination and the significant ones were kept using *envfit* fuction in *vegan* package. Permutational Multivariate Analysis of Variance (PERMANOVA) was conducted to test the significance of impacts of population density and greenspace type on the microbial community composition. The importance of the environmental factors for community composition was determined with multiple regressions on matrices (MRM) using *ecodist* package. The relationship between Bray-Curtis dissimilarities and geographic distances were estimated using linear regression. Sub-networks for each site were generated from the meta-community network by conserving all the MED nodes existing in samples from each site using *subgraph* function and several network-level topological features (SI Table S1) were estimated in *igraph* package. Kruskal-Wallis post-hoc tests were employed to compare the means of alpha diversity, soil properties and topological features between different groups of a factor using *PMCMR* package. All the *P*-values for multiple comparisons were adjusted by false discovery rate (FDR) method and the null hypothesis was rejected while *P*-values were less than 0.05.

Results

Soil properties

We collected 5 replicate soil samples from 150 sites across Chicago for 16S rRNA amplicon sequencing, but sub-samples of each replicate from the same site were pooled. In total, 150 pooled samples representing 150 sites were sent for physicochemical analysis. Of the measured soil edaphic properties, only pH and moisture were significantly different across population densities and greenspace types (SI Figure S3 and S4). Soil pH and moisture were both significantly greater in high population density areas (Kruskal-Wallis post-hoc $P<0.05$) (SI Figure S3). In addition, pH was greater in soils proximal to a two-way street, compared to soils in one-way streets (Kruskal-Wallis post-hoc $P<0.05$), large parks (Kruskal-Wallis post-hoc $P<0.01$) and residential areas (Kruskal-Wallis post-hoc $P<0.01$) (SI Figure S4); meanwhile soil moisture was significant greater in park soils (Kruskal-Wallis post-hoc $P<0.001$) and residential areas (Kruskal-Wallis post-hoc $P<0.05$) compare to street soils (SI Figure S4).

Microbial community analysis

All 750 soil samples were processed for 16S rRNA amplicon sequencing, generating 12,387,819 sequences. These were processed using Minimal Entropy Decomposition to produce 37,927 nodes. The MED node matrix was rarefied to 8,000 sequences per sample (SI Figure S2), resulting in only 719 samples being used in subsequent analysis. Proteobacteria (30.2%), Acidobacteria (20.7%), Bacteroidetes (17.2%), Verrucomicrobia (11.0%) and Actinobacteria (6.0%) dominated the sampling sites. Within these phyla, the most abundant classes were the 6th subdivision of the Acidobacteria (11.0%), Saprospirae (9.9%) and Alphaproteobacteria (8.9%) (SI Figure S5). The most abundant genera included unknown genera belonging to the order iii1-15 (8.5%), and the families Syntrophobacteraceae (3.5%), Ellin6075 (3.1%), and Chitinophagaceae (7.6%). The genera *DA101* (4.6%; potentially *Candidatus* Udaeobacter copiosus) and *Candidatus*

234 Nitrososphaera (3.2%) were also abundant (SI Figure S6). The majority of taxa were common to
235 both population densities, and all greenspace types (SI Figure S7).
236 Population density was significantly positively correlated with both Shannon and phylogenetic
237 diversity (Spearman's rank $P<0.001$) (Table 1). Greenspace type also correlated with both
238 phylogenetic and Shannon's diversity (Spearman's rank $P<0.05$; Table 1). We divided population
239 density into 5 categories (density1-5), and the average Shannon diversity was greater in each
240 successive density category. However, the phylogenetic diversity was lower in density4 compared
241 to density3 and density5 (Figure 1a and b). The small park and one-way street soils showed both
242 higher Shannon and phylogenetic diversities than the other soils (Figure 1c and 1d). Meanwhile,
243 park soils had significantly greater overall diversity than residential soils, while soils proximal to
244 a street had greater Shannon diversity than residential soils, but showed no significant difference
245 in phylogenetic diversity (Figure 1c and 1d). Moisture, nitrate and the proportions of clay and silt,
246 were negatively correlated with microbial diversity while the proportion of sand in the soil
247 positively correlated with the diversity (Table 1). We used the Krige method of data interpolation
248 to predict the distribution of microbial diversity in regions encompassing the City of Chicago
249 (Figure 2). Predicted microbial diversity was greater in the north than in the south, which correlated
250 with the distribution of population density (SI Figure S1). We calculated the statistical power to
251 demonstrate differences in the microbiota between greenspace type and population density, with
252 different sample sizes based on Shannon index (SI Figure S8). At an 80% power level, 75 and 115
253 samples were needed to obtain the difference between small park and residential area at significant
254 levels of 0.05 and 0.01, respectively; 53 and 80 samples were needed to obtain the difference
255 between density5 and density1 at significant levels of 0.05 and 0.01, respectively.

Microbial community composition was significantly associated with population density (PERMANOVA, R^2=0.007, *P*=0.001) and greenspace type (PERMANOVA, R^2=0.039, *P*=0.001). The moisture, pH and soil texture (sand, silt, or clay) described the largest degree of variance in microbial community composition (Figure 3). The soil moisture and texture differentiated the park and street soil microbial communities while pH differentiated the communities from residential and street soils (Figure 3). We applied a multiple regression on matrices (MRM) model, which confirmed that population density, greenspace type, and soil physicochemistry (except C/N ratio, nitrate concentration, and percentage of clay and silt) significantly associated with microbial community structure (SI Table S2). However, the most significant driver was pH, which may be explaining the majority of variance seen with changes in population density and greenspace type. Microbial community dissimilarity and geographic distance were also significantly and positively correlated with each other (linear regression, R^2=0.241, *P*=0.002; SI Figure S9), which again is likely driven by changes in pH and potentially soil moisture.

LEfSe analysis was employed to determine the taxa that differentiated between categorical variables (Figure 4 and SI Figure S10). Nitrospirae, Acidobacteria and Verrucomicrobia, the family Piscirickettsiaceae and genus *Candidatus Xiphinematobacter* were significantly enriched in areas with low population density, while Alphaproteobacteria, Bacteroidetes, Firmicutes, Gemmatimonadetes and Chloroflexi, including families Chitinophagaceae and Sphingobacteriaceae were enriched in high density areas (Figure 4a and SI Figure S10a). Deltaproteobacteria, including *Geobacter*, auto67-4W and Bradyrhizobiaceae were significantly more abundant in parkland soils. While soils proximal to streets were dominated by Chitinophagaceae, Sphingobacteriaceae, *Steroidobacter*, *Cellvibrio*, Gemmatimonadetes, and

Ellin6075. The genera *DA101* (likely *Candidatus* Udaeobacter copiosus) and *Nitrospria* were among the dominant taxa in residential zones (Figure 4b and SI Figure S10b).

Topological features of the co-occurrence network

A co-occurrence network was constructed based on correlation relationships with a deconvolution procedure. In this network, 92,670 significant correlations were captured among 1,502 MED nodes (SI Figure S11). These correlations could indicate the potential interactions among different microbial taxa. Each MED node was shown as a vertex in the network. The distribution of degrees showed that most of the vertices were with a degree between 100 and 200 (SI Figure S12a). The relative abundance of the vertex was significantly and positively correlated with the degree, closeness and betweenness centrality and negatively correlated with transitivity (Figure 5a and SI Figure S12b), suggesting that taxa with the greatest abundance had greater interaction. Negative correlations between density ratio and topological features including degree, closeness centrality and transitivity indicated that the abundant soil microbes in high density population areas had fewer interactions, compared to low population density areas (Figure 5a). Meanwhile moisture ratio was positively correlated with topological features including degree, closeness centrality, betweenness centrality and transitivity, suggesting a greater interaction of the abundant taxa in high moisture soils, compared to low moisture soils (Figure 5a). While maintaining the vertices from each site, sub-networks were constructed for each site and network-level topological features were calculated. pH was negatively correlated with edge density and assortativity degree (Figure 5b). Soil moisture was positively correlated with average closeness, edge density and assortativity degree, while it was negatively correlated with average path (Figure 5b), suggesting a closer relationship of the microbes in soils with a higher water content. For soil texture, sand content showed negative correlations with average closeness, edge density and assortativity degree but

positive correlation with average path (Figure 5b). However, clay and silt exhibited the opposite trend, indicating that larger sized particles leaded to a weaker interaction of the microbes in soils. Population density was negatively correlated with average closeness and edge density, and positive relationships with average betweenness, vertex number and edge number were also observed (Figure 5b). The average betweenness was significantly higher in the density5 soils compared to density1 soils, while vertex number showed a contrast trend (SI Figure S13). The density decreased with the increasing population densities (SI Figure S13). The average path length was significantly greater in two-way street soils while the other network features (except average betweenness and diameter) were lower in these soil samples (SI Figure S14), which suggested simpler structures for these networks. The shorter average path length in park soils compared to street and residential soils implied a greater degree of association between changes in abundance for taxa within park soils (SI Figure S14).

Discussion

We characterized the bacterial diversity, community composition and interactions between taxa in soils from across regions of varying population density and greenspace types in a major urban environment. This analysis demonstrated that while moisture was the key driver of microbial community structure across the region, this is likely being driven by varying management practice and pollution impact that associate with varying population density and greenspace type. The human population density was positively correlated with microbial diversity. The private residential soils maintained lower microbial diversity compared to the public street and park soils. All the microbial taxa in soils were well interacted although network topology differences existed between different types of soils and population densities. These findings hint that anthropogenic

activities and managements of the turf-grasses associated with the changes observed in this study might significantly influence the underground communities, which may contribute to environmental issues through perturbing the nutrient cycles.

The relationship between microbial diversity and environmental function has received substantial attention. It is recognized that loss of microbial diversity is negatively associated with ecosystem function and can also be associated with human health problems.[32-36] We found a significant and positive correlation between population density and microbial diversity. This congruence between people and microbial diversity is in agreement with other studies focusing on plants, mammals and birds, which indicate spatial congruence as a function of energy availability, land transformation, disturbance, and economic factors.[15] We observed higher bacteria diversity in more densely populated urban areas as compared to low-density areas. Moreover, we found that microbial diversity was greater in the park and street soils than in the residential soils. This may be partly due to the differences in soil moisture, which also significantly correlated with diversity, and was significantly different between different greenspace soils. Interestingly, sand content was found to be positively correlated with bacterial diversity while clay and silt showed the opposite trend, indicating that larger sized particles favored more species. It is suggested that a coarser soil would possess more isolated water films that could create more isolated microhabitats to harbor microbial species.[37] This could lead to higher diversity in the soils with more sand. In addition, the effect of habitat size may contribute to the observed results since urban patch area has positive effects on biodiversity.[38] In contrast, a study in New York City (NY, USA) found no difference of bacterial diversity between small (street median) and large habitats (parks), although the fungal richness was lower in street median compare to the parks.[6] Another study conducted in Xiamen (China) showed no difference of bacterial diversity between urban and suburban turfgrass soils

that were associated with different human population densities.[39] However, these studies only investigated a few certain sites in a city, while our study sampled across the whole city with an appropriate statistical power, based on sample size calculations for detecting variance. These varied results also reveal that other factors in addition to population density and habitat type can influence microbial diversity. Therefore, future studies on this subject should consider more site features including history of soil management.

To examine the microbial diversity distribution at a whole city scale, we predicted the distribution of Shannon and phylogenetic diversity. Interestingly, a significant separation was found between the north and south Chicago. This pattern of distribution was similar to that of the population density. However, some exceptions were found. For instance, the northeast and middle-east had lower population densities but higher microbial diversities. These sites were closely associated with major international airports (Chicago O'Hare and Midway), which could be interpreted as urban hotspots of activity reflecting microbial hotspots. It is suggested that turning farmlands into urban areas associates with a decrease in total microbial diversity, which may imperil the soil functions.[39] This study reveals that within urban areas, population density positively associated with microbial diversity, but no convincible conclusion can be made since the diversity shifted within a small range. However, it is worth noting that this significant association might be used as a potential evaluation of our living space since the decline of microbial diversity might lead to some deseases and public health issues.[35,36,40]

While the core microbiota were mostly shared between soils associated with different population densities and greenspace types, the bacterial community structure did differ between high and low densities as well as between different greenspaces. The significant impact of soil texture on the bacterial community composition suggests that paricle size, as the proxy of soil type, is crutial to

the structure of the microbial residents. We also found that the soil moisture was the most important factor that shaped the community structure. It has been suggested that habitats in the city associated with street medians with less plant canopy, are hotter, drier and surrounded by more impervious surfaces compared to the parks.[41] Similarly, as revealed in this study, the moisture was significantly greater in park soils than in street soils, which could explain the abundance of anaerobic taxa such as *Geobacter* in moisture rich park soils,[42,43] taxa adapted to low moisture content such as Gemmatimonadetes dominated in street soils.[44] The greater abundance of *Nitrospria* in residential areas, suggests that the private lawns likely receive greater nitrogen fertilizer than other soils, which would promote the growth of nitrifying microbes. Interestingly, taxa associated with the Verrucomicrobia, which are generally slow-growing and oligotrophic[45,46] were dominated in low population density areas, suggesting that these area might be both nitrogen poor and carbon rich, as many Verrucomicrobia are specialized in the degradation of recalcitrant carbon compounds.[47] These results indicate that improper management and potential pollutions associated with the gradient of population density and greenspace type might contribute to the shifts in microbial communities, thereby influencing the soil health.

The effects of geographic distances on the microbial diversity and community structure have been broadly studied at the continental scale.[48-51] However, the information on this topic at a smaller scale remains limited. Here we found that the geographic distances were significantly correlated with dissimilarities of microbial community at a whole city scale. Given that association between distances and community dissimilarities implies the influence of historical contingencies which were considered the key factor driving the microbial biogeography,[52] our study suggests that multiple land scales should be included in future works to thoroughly understand the evolutionary characteristics of microorganisms.

Assessing the potential interactions of microbes using network inference could help with understanding the complex microbial communities and building models of ecosystem dynamics.[53] In our study, microbes in urban greenspaces were well-interacted shown as numerous co-occurrence correlations. The abundant taxa in the soils tended to have greater degree and centrality values, while their neighbors were generally less connected. One explanation for this trend could be that the abundant taxa maintained core niches in these urban soils upon which other taxa were either dependent, or they were co-dependent on a shared resource. We found that the microbial taxa were more interacted in low population density soils, which correlates with reduced microbial diversity, while in two-way street soils, the microbial taxa were less interacted. These results could be explained by the varied soil moisture since there were positive correlations between moisture and topological features which indicate that greater moisture content leads to more interactions of the microorganisms. This supports previous observations that soil moisture has a strong impact on microbial community structure. One explanation is that high water content conditions in soils create greater homogeneity, and hence weak niche differentiation which contributes to stronger interactions between soils microbes.[53]

Given that urban habitats are undergoing rapid population expansion, and associated changes in land-use,[1] improving our knowledge of the impact of these changes on biodiversity is essential. While a few studies have hinted at an association between microbial biodiversity and population density, this study is the first to comprehensively confirm this association. In addition, microbial diversity was significantly differentiated between different greenspace types. The importance of soil moisture, pH and texture in shaping microbial community structure and co-occurrence patterns was also confirmed. These results might inform future urban planning and land management practice for urban green infrastructure, thereby improving the quality of these environments.

Despite this, a full understanding of the relationship between urbanization, biodiversity and ecosystem function remains a key goal, and must be conducted at multiple scales with appropriate statistical power.

Acknowledgement

This study is supported by the Earth Microbiome Project (http://www.earthmicrobiome.org/) and the China Scholarship Council (http://en.csc.edu.cn/). We thank Margaret Sladek for her help with the sampling and Neil Gottel for his help with sample processing. We thank Cesar Cardona, Bin Ma and Jianming Xu for their help with data analysis.

Supporting Information

Additional figures related to the distribution of sampling sites, rarefaction curves, soil properties in different categories, distributions of abundant taxa and core microbiome, the statistical power associated with sample size, the regression between microbial dissimilarities and geographic distances, LDA scores for differently abundant taxa, and the information of co-occurrence network and topological features; additional tables showing the meaning of topological features, and results of MRM model.

References

(1) Grimm, N. B.; Faeth, S. H.; Golubiewski, N. E.; Redman, C. L.; Wu, J.; Bai, X.; Briggs, J. M. Global change and the ecology of cities. *Science* **2008,** *319* (5864), 756-760.

(2) Zhu, Y. G.; Ioannidis, J. P.; Li, H.; Jones, K. C.; Martin, F. L. Understanding and harnessing the health effects of rapid urbanization in China. *Environ. Sci. Technol.* **2011,** *45* (12), 5099-5104.

(3) Falkowski, P. G.; Fenchel, T.; Delong, E. F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **2008,** *320* (5879), 1034-1039.

(4) Kaye, J.; McCulley, R.; Burke, I. Carbon fluxes, nitrogen cycling, and soil microbial communities in adjacent urban, native and agricultural ecosystems. *Glob. Change Biol.* **2005,** *11* (4), 575-587.

(5) Yan, B.; Li, J.; Xiao, N.; Qi, Y.; Fu, G.; Liu, G.; Qiao, M. Urban-development-induced Changes in the Diversity and Composition of the Soil Bacterial Community in Beijing. *Sci. Rep.* **2016,** *6*, 38811.

(6) Reese, A. T.; Savage, A.; Youngsteadt, E.; McGuire, K. L.; Koling, A.; Watkins, O.; Frank, S. D.; Dunn, R. R. Urban stress is associated with variation in microbial species composition-but not richness-in Manhattan. *ISME J.* **2016,** *10* (3), 751-760.

(7) Groffman, P. M.; Williams, C. O.; Pouyat, R. V.; Band, L. E.; Yesilonis, I. D. Nitrate leaching and nitrous oxide flux in urban forests and grasslands. *J. Environ. Qual.* **2009,** *38* (5), 1848-1860.

(8) Milesi, C.; Running, S. W.; Elvidge, C. D.; Dietz, J. B.; Tuttle, B. T.; Nemani, R. R. Mapping and modeling the biogeochemical cycling of turf grasses in the United States. *Environ. Manage.* **2005,** *36* (3), 426-438.

(9) King, G. M. Urban microbiomes and urban ecology: how do microbes in the built environment affect human sustainability in cities? *J. Microbiol.* **2014,** *52* (9), 721-728.

(10) Groffman, P. M.; Boulware, N. J.; Zipperer, W. C.; Pouyat, R. V.; Band, L. E.; Colosimo, M. F. Soil nitrogen cycle processes in urban riparian zones. *Environ. Sci. Technol.* **2002,** *36* (21), 4547-4552.

(11) Charlop-Powers, Z.; Pregitzer, C. C.; Lemetre, C.; Ternei, M. A.; Maniko, J.; Hover, B. M.; Calle, P. Y.; McGuire, K. L.; Garbarino, J.; Forgione, H. M.; Charlop-Powers, S.; Brady, S. F.

Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proc. Natl. Acad. Sci. USA* **2016,** *113* (51), 14811-14816.

(12) Ramirez, K. S.; Leff, J. W.; Barberan, A.; Bates, S. T.; Betley, J.; Crowther, T. W.; Kelly, E. F.; Oldfield, E. E.; Shaw, E. A.; Steenbock, C.; Bradford, M. A.; Wall, D. H.; Fierer, N. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. Biol. Sci.* **2014,** *281* (1795), 22.

(13) Robbins, P.; Sharp, J. T. Producing and consuming chemicals: the moral economy of the American lawn. *Econ. Geogr.* **2003,** *79* (4), 425-451.

(14) Jenkins, V., *The lawn: a history of an American obsession*. Smithsonian Institution: 2015.

(15) Luck, G. W. A review of the relationships between human population density and biodiversity. *Biol Rev Camb Philos Soc* **2007,** *82* (4), 607-645.

(16) Luck, G. W. The relationships between net primary productivity, human population density and species conservation. *J. Biogeogr.* **2007,** *34* (2), 201-212.

(17) Luck, G. W.; Ricketts, T. H.; Daily, G. C.; Imhoff, M. Alleviating spatial conflict between people and biodiversity. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (1), 182-186.

(18) Chown, S. L.; van Rensburg, B. J.; Gaston, K. J.; Rodrigues, A. S.; van Jaarsveld, A. S. Energy, species richness, and human population size: conservation implications at a national scale. *Ecol. Appl.* **2003,** *13* (5), 1233-1241.

(19) Balmford, A.; Moore, J. L.; Brooks, T.; Burgess, N.; Hansen, L. A.; Williams, P.; Rahbek, C. Conservation conflicts across Africa. *Science* **2001,** *291* (5513), 2616-2619.

(20) Lomolino, M. V.; Smith, G. A. Terrestrial vertebrate communities at black-tailed prairie dog (*Cynomys ludovicianus*) towns. *Biol. Conserv.* **2004,** *115* (1), 89-100.

(21) Woodroffe, R. Predators and people: using human densities to interpret declines of large carnivores. *Anim. Conserv.* **2000,** *3* (2), 165-173.

(22) Bachrach, J. S., *The City in a Garden: A Photographic History of Chicago's Parks*. Center for American Places, 2001.

(23) Ma, B.; Wang, H.; Dsouza, M.; Lou, J.; He, Y.; Dai, Z.; Brookes, P. C.; Xu, J.; Gilbert, J. A. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* **2016,** *10* (8), 1891-1901.

(24) Walters, W.; Hyde, E. R.; Berg-Lyons, D.; Ackermann, G.; Humphrey, G.; Parada, A.; Gilbert, J. A.; Jansson, J. K.; Caporaso, J. G.; Fuhrman, J. A. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* **2016,** *1* (1), e00009-15.

(25) Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010,** *7* (5), 335-336.

(26) Eren, A. M.; Morrison, H. G.; Lescault, P. J.; Reveillaud, J.; Vineis, J. H.; Sogin, M. L. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **2015,** *9* (4), 968-979.

(27) Edgar, R. C.; Haas, B. J.; Clemente, J. C.; Quince, C.; Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **2011,** *27* (16), 2194-2200.

(28) McDonald, D.; Price, M. N.; Goodrich, J.; Nawrocki, E. P.; DeSantis, T. Z.; Probst, A.; Andersen, G. L.; Knight, R.; Hugenholtz, P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **2012,** *6* (3), 610-618.

(29) Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W. S.; Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol.* **2011,** *12* (6), R60.

(30) Doerks, T.; Copley, R. R.; Schultz, J.; Ponting, C. P.; Bork, P. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Re.s* **2002,** *12* (1), 47-56.

(31) Chow, S.-C.; Wang, H.; Shao, J., *Sample size calculations in clinical research*. CRC press, 2007.

(32) Heederik, D.; von Mutius, E. Does diversity of environmental microbial exposure matter for the occurrence of allergy and asthma? *J. Allergy Clin. Immunol.* **2012,** *130* (1), 44-50.

(33) Singh, B. K.; Quince, C.; Macdonald, C. A.; Khachane, A.; Thomas, N.; Al-Soud, W. A.; Sorensen, S. J.; He, Z.; White, D.; Sinclair, A.; Crooks, B.; Zhou, J.; Campbell, C. D. Loss of microbial diversity in soils is coincident with reductions in some specialized functions. *Environ. Microbiol.* **2014,** *16* (8), 2408-2420.

(34) Philippot, L.; Spor, A.; Henault, C.; Bru, D.; Bizouard, F.; Jones, C. M.; Sarr, A.; Maron, P. A. Loss in microbial diversity affects nitrogen cycling in soil. *ISME J.* **2013,** *7* (8), 1609-1619.

(35) Rook, G. A. Review series on helminths, immune modulation and the hygiene hypothesis: the broader implications of the hygiene hypothesis. *Immunology* **2009,** *126* (1), 3-11.

(36) Ege, M. J.; Mayer, M.; Normand, A.-C.; Genuneit, J.; Cookson, W. O.; Braun-Fahrländer, C.; Heederik, D.; Piarroux, R.; von Mutius, E. Exposure to environmental microorganisms and childhood asthma. *New Engl. J. Med.* **2011,** *364* (8), 701-709.

(37) Chau, J. F.; Bagtzoglou, A. C.; Willig, M. R. The Effect of Soil Texture on Richness and Diversity of Bacterial Communities. *Environ. Forensics* **2011,** *12* (4), 333-341.

(38) Beninde, J.; Veith, M.; Hochkirch, A. Biodiversity in cities needs space: a meta-analysis of factors determining intra-urban biodiversity variation. *Ecol. Lett.* **2015,** *18* (6), 581-592.

(39) Wang, H. T.; Marshall, C. W.; Cheng, M. Y.; Xu, H. J.; Li, H.; Yang, X. R.; Zheng, T. L. Changes in land use driven by urbanization impact nitrogen cycling and the microbial community composition in soils. *Sci. Rep.* **2017,** *7*, 44049.

(40) Hanski, I.; von Hertzen, L.; Fyhrquist, N.; Koskinen, K.; Torppa, K.; Laatikainen, T.; Karisola, P.; Auvinen, P.; Paulin, L.; Mäkelä, M. J. Environmental biodiversity, human microbiota, and allergy are interrelated. *Proc. Natl. Acad. Sci. USA* **2012,** *109* (21), 8334-8339.

(41) Youngsteadt, E.; Henderson, R. C.; Savage, A. M.; Ernst, A. F.; Dunn, R. R.; Frank, S. D. Habitat and species identity, not diversity, predict the extent of refuse consumption by urban arthropods. *Glob. Change Biol.* **2015,** *21* (3), 1103-1115.

(42) Lovley, D. R.; Stolz, J. F.; Nord, G. L.; Phillips, E. J. Anaerobic production of magnetite by a dissimilatory iron-reducing microorganism. *Nature* **1987,** *330* (6145), 252-254.

(43) Coates, J. D.; Phillips, E.; Lonergan, D. J.; Jenter, H.; Lovley, D. R. Isolation of Geobacter species from diverse sedimentary environments. *Appl. Environ. Microbiol.* **1996,** *62* (5), 1531-1536.

(44) DeBruyn, J. M.; Nixon, L. T.; Fawaz, M. N.; Johnson, A. M.; Radosevich, M. Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. *Appl. Environ. Microbiol.* **2011,** *77* (17), 6295-6300.

(45) Senechkin, I. V.; Speksnijder, A. G.; Semenov, A. M.; Van Bruggen, A. H.; Van Overbeek, L. S. Isolation and partial characterization of bacterial strains on low organic carbon medium from soils fertilized with different organic amendments. *Microb. Ecol.* **2010,** *60* (4), 829-839.

(46) Janssen, P. H.; Yates, P. S.; Grinton, B. E.; Taylor, P. M.; Sait, M. Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Appl. Environ. Microbiol.* **2002,** *68* (5), 2391-2396.

(47) Herlemann, D. P.; Lundin, D.; Labrenz, M.; Jurgens, K.; Zheng, Z.; Aspeborg, H.; Andersson, A. F. Metagenomic de novo assembly of an aquatic representative of the verrucomicrobial class Spartobacteria. *MBio* **2013,** *4* (3), e00569-12.

(48) Ma, B.; Dai, Z.; Wang, H.; Dsouza, M.; Liu, X.; He, Y.; Wu, J.; Rodrigues, J. L.; Gilbert, J. A.; Brookes, P. C.; Xu, J. Distinct Biogeographic Patterns for Archaea, Bacteria, and Fungi along the Vegetation Gradient at the Continental Scale in Eastern China. *mSystems* **2017,** *2* (1), e00174-16.

(49) Fierer, N.; Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (3), 626-631.

(50) Ge, Y.; He, J.-z.; Zhu, Y.-g.; Zhang, J.-b.; Xu, Z.; Zhang, L.-m.; Zheng, Y.-m. Differences in soil bacterial diversity: driven by contemporary disturbances or historical contingencies? *ISME J.* **2008,** *2* (3), 254-264.

(51) Xiong, J.; Liu, Y.; Lin, X.; Zhang, H.; Zeng, J.; Hou, J.; Yang, Y.; Yao, T.; Knight, R.; Chu, H. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ. Microbiol.* **2012,** *14* (9), 2457-2466.

(52) Martiny, J. B.; Bohannan, B. J.; Brown, J. H.; Colwell, R. K.; Fuhrman, J. A.; Green, J. L.; Horner-Devine, M. C.; Kane, M.; Krumins, J. A.; Kuske, C. R.; Morin, P. J.; Naeem, S.; Ovreas, L.; Reysenbach, A. L.; Smith, V. H.; Staley, J. T. Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **2006,** *4* (2), 102-112.

(53) Faust, K.; Raes, J. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **2012,** *10* (8), 538-550.

Figure captions

Figure 1 The boxplots showing the distribution of Shannon index and phylogenetic diversity by different population density levels (a and b) or different greenspace types (c and d). Differences are significant when no same letter exists between groups ($P<0.05$). The darkred diamonds indicate the mean values. *, $P<0.05$; **, $P<0.01$.

Figure 2 Spatial mapping of Shannon index (a) and phylogenetic diversity (b) across the City of Chicago using kriging interpolation.

Figure 3 nMDS plot based on the Bray-Cutis dissimilarities. The main ordination shows similarity between samples and the arrows show correlations between environmental variables and ordination axes. The centroids representing different types of greenspace categorical variable are shown as corresponding diamonds.

Figure 4 Strict version of LEfSe results on microbial communities. The cladogram indicates the taxa (highlighted with small circles and shading) showing different abundance values (according to LEfSe) in high and low population density soils (a) or in park, street and residential soils (b). For each taxon (circle), the color denotes the significantly higher abundance of the taxon in the corresponding group. Yellow denotes that the taxon is not significantly higher in any group.

Figure 5 The Spearman's correlation analysis between soil properties, human population density, greenspace type and vertex-level (a) or network-level (b) topological features. All the significant correlations are shown ($P<0.05$). The five greenspace types were transformed to dummy variables (Presence, 1; Absence, 0). Abundance, the relative abundances of each MED node; Density ratio:

the average abundance of the MED node in high population density samples to that in low population density samples; pH ratio, the average abundance of the MED node in high pH (above the average) samples to that in low pH (below the average) samples; Moisture ratio, the average abundance of the MED node in high moisture (above the average) samples to that in low moisture (below the average) samples. TC, total carbon; TN, total nitrogen; OM, organic matter.

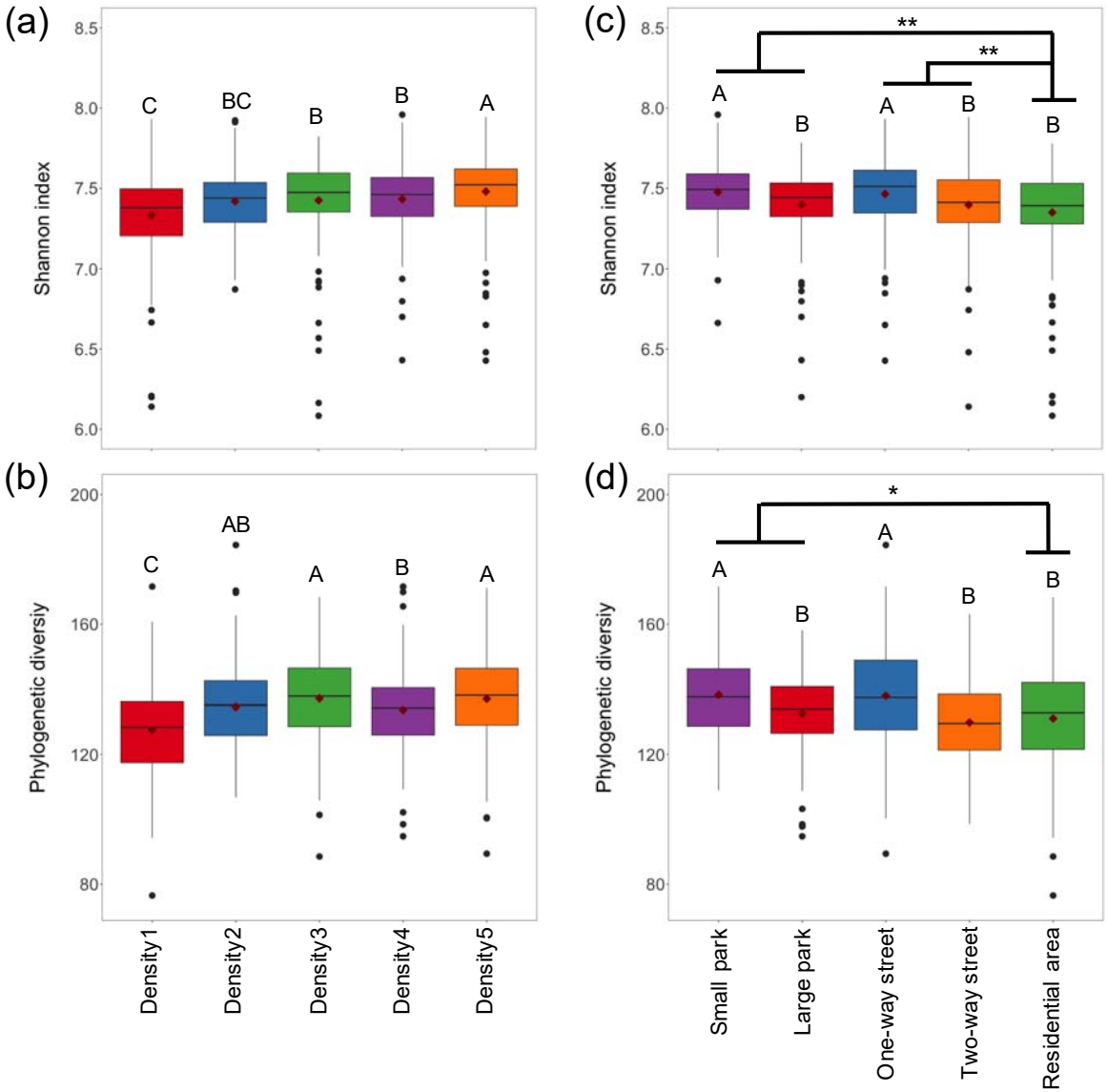
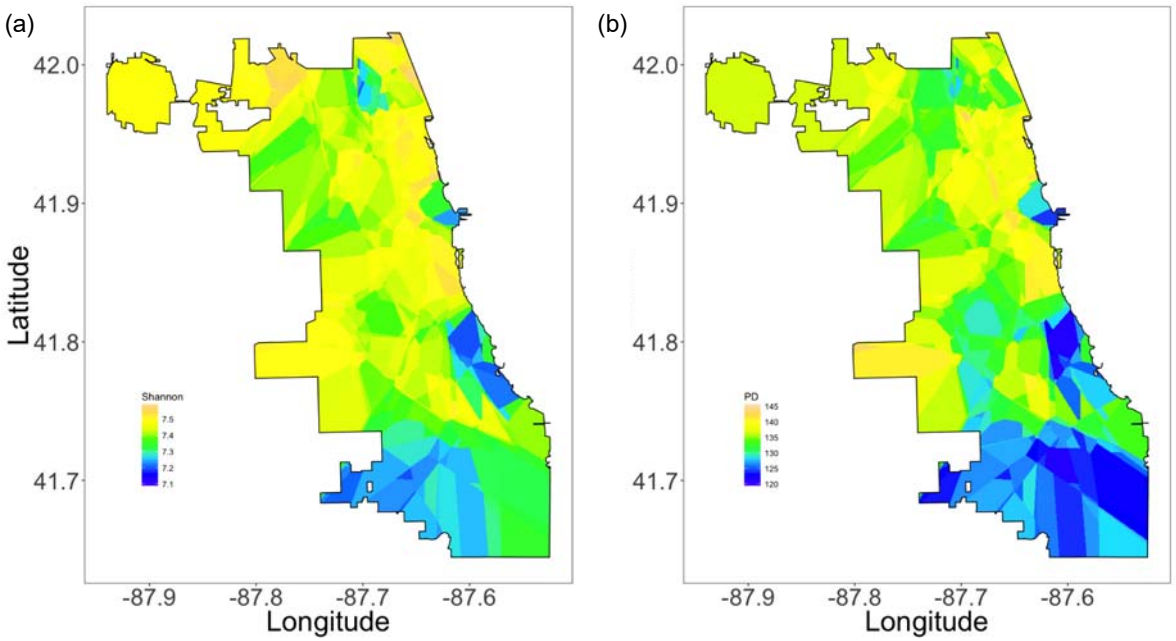
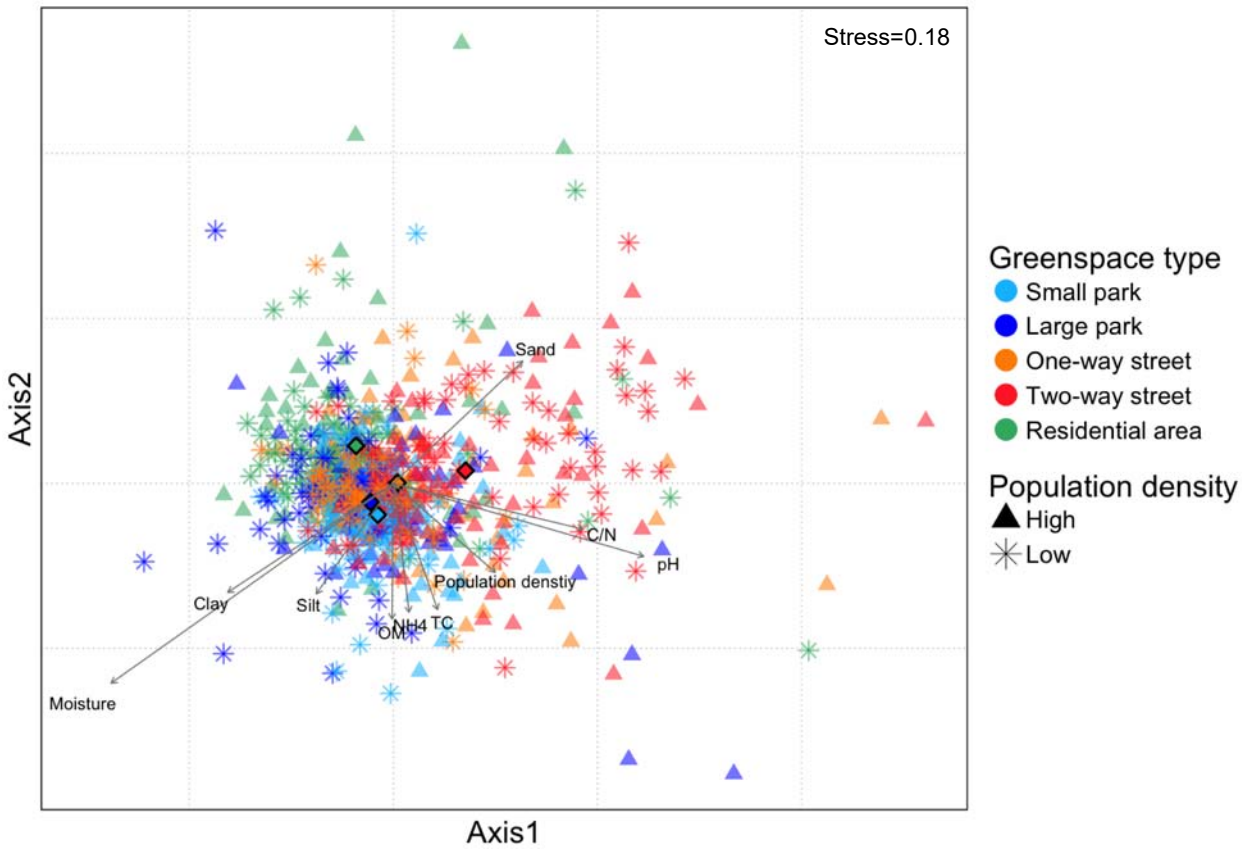

(a)
Shannon index
8.5
8.0
7.5
7.0
6.5
6.0
C
BC
B
B
A
(b)
Phylogenetic diversiy
200
160
120
80
C
AB
A
B
A
Density1
Density2
Density3
Density4
Density5
(c)
Shannon index
**
**
A
B
A
B
B
(d)
Phylogenetic diversiy
*
A
B
A
B
B
Small park
Large park
One-way street
Two-way street
Residential area

Figure 1



607
608

Figure 2



609
610

Figure 3

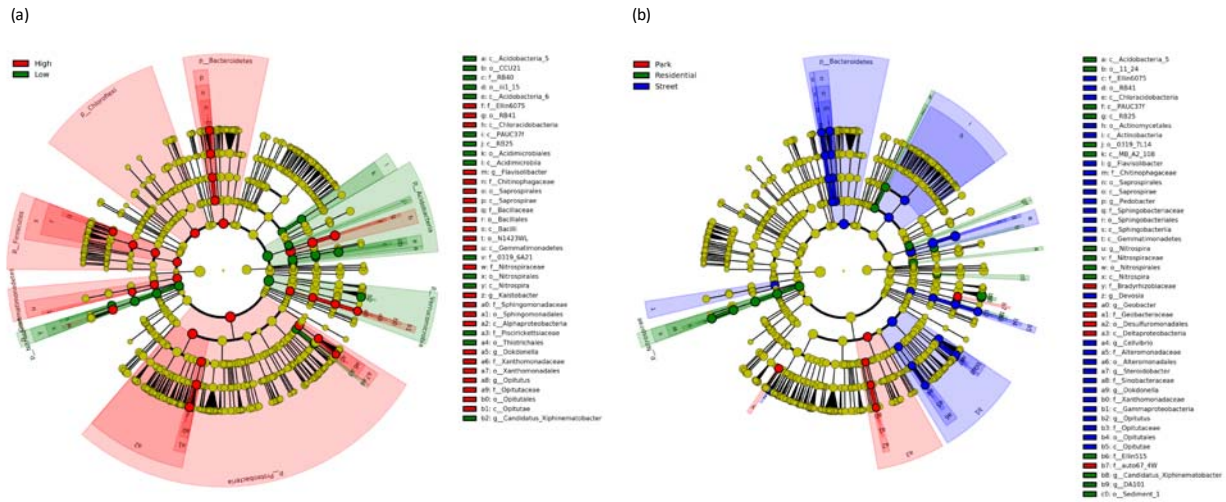


Figure 4

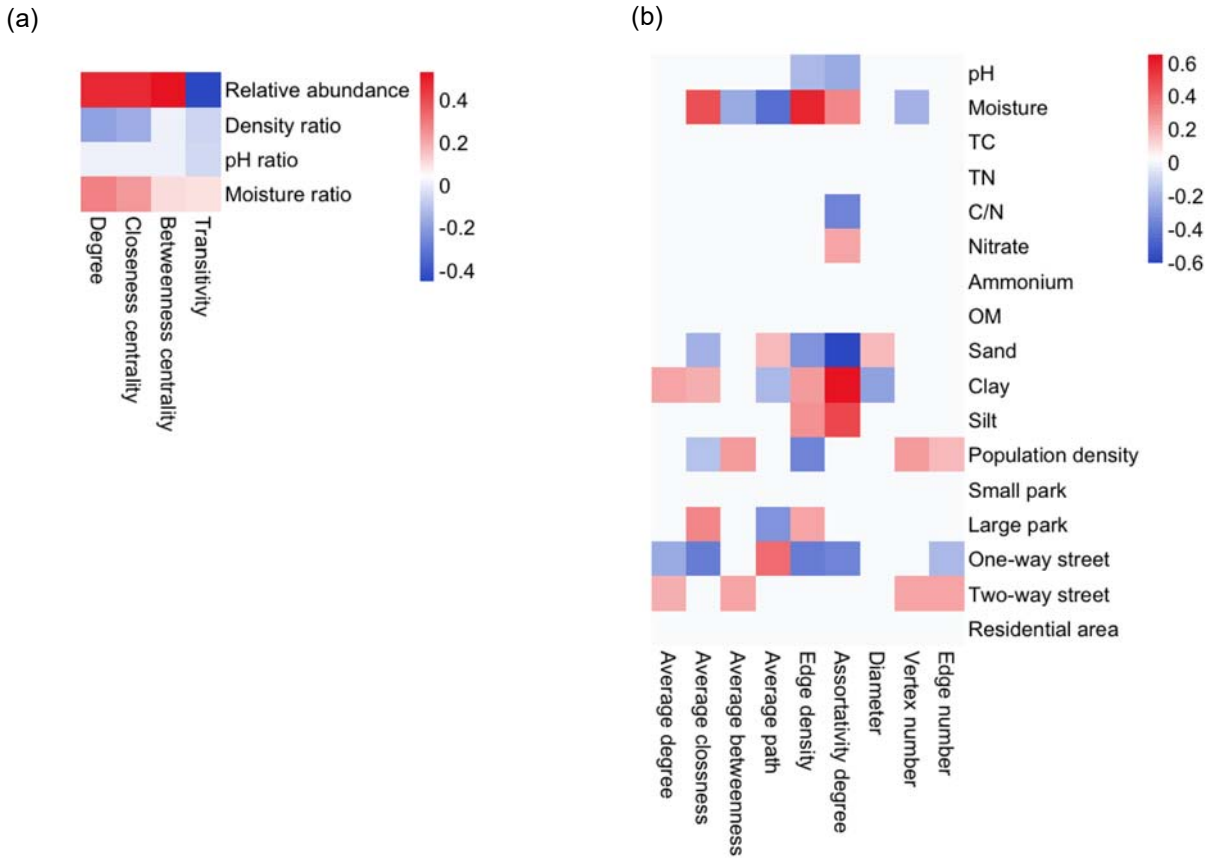


Figure 5

Table1 The Spearman's rank correlation analysis between soil properties, human population density, greenspace types and alpha diversity

	Shannon index		Phylogenetic diversity	
	Coefficients	*P* value	Coefficients	*P* value
pH	0.080	0.875	0.052	0.965
Moisture	-0.237	0.014*	-0.236	0.014*
TC	0.012	0.886	-0.130	0.230
TN	-0.053	0.868	-0.145	0.612
C/N ratio	0.122	0.363	0.095	0.391
Nitrate	-0.203	0.031*	-0.208	0.031*
Ammonium	-0.082	0.589	-0.098	0.589
OM	-0.050	0.555	-0.149	0.436
Sand	0.346	<0.001***	0.198	0.024*
Clay	-0.305	<0.001***	-0.132	0.173
Silt	-0.316	<0.001***	-0.226	0.011*
Population density	0.236	0.009**	0.270	0.003**
Small park	0.164	0.061	0.223	0.049*
Large park	-0.104	0.330	-0.079	0.385
One-way street	-0.080	0.377	-0.228	0.010*
Two-way street	0.144	0.092	0.211	0.015*
Residential area	-0.124	0.262	-0.128	0.262

The five greenspace types were transformed to dummy variables (Presence, 1; Absence, 0). All the *P* values were adjusted with FDR method. TC, total carbon; TN, total nitrogen; OM, organic matter; *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$.