PROCEEDINGS OF THE ROYAL SOCIETY B | BIOLOGICAL SCIENCES

Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally

Kelly S. Ramirez, Jonathan W. Leff, Albert Barberán, Scott Thomas Bates, Jason Betley, Thomas W. Crowther, Eugene F. Kelly, Emily E. Oldfield, E. Ashley Shaw, Christopher Steenbock, Mark A. Bradford, Diana H. Wall and Noah Fierer

Supplementary data	"Data Supplement" http://rspb.royalsocietypublishing.org/content/suppl/2014/09/30/rspb.2014.1988.DC1.html
References	**This article cites 49 articles, 16 of which can be accessed free** http://rspb.royalsocietypublishing.org/content/281/1795/20141988.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections ecology (1764 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

PROCEEDINGS OF THE ROYAL SOCIETY B

Research

CrossMark
click for updates

Author for correspondence:
Noah Fierer
e-mail: noah.fierer@colorado.edu

Royal Society Publishing

Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally

Kelly S. Ramirez[1], Jonathan W. Leff[4,5], Albert Barberán[5], Scott Thomas Bates[6], Jason Betley[7], Thomas W. Crowther[8], Eugene F. Kelly[2], Emily E. Oldfield[8], E. Ashley Shaw[3], Christopher Steenbock[4], Mark A. Bradford[8], Diana H. Wall[1,3] and Noah Fierer[4,5]

[1]School of Global Environmental Sustainability, [2]Department of Soil and Crop Sciences, and [3]Department of Biology, Colorado State University, Fort Collins, CO 80523, USA
[4]Department of Ecology and Evolutionary Biology, and [5]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA
[6]Department of Plant Pathology, University of Minnesota, Saint Paul, MN 55108, USA
[7]Illumina UK, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL, UK
[8]School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511, USA

Soil biota play key roles in the functioning of terrestrial ecosystems, however, compared to our knowledge of above-ground plant and animal diversity, the biodiversity found in soils remains largely uncharacterized. Here, we present an assessment of soil biodiversity and biogeographic patterns across Central Park in New York City that spanned all three domains of life, demonstrating that even an urban, managed system harbours large amounts of undescribed soil biodiversity. Despite high variability across the Park, below-ground diversity patterns were predictable based on soil characteristics, with prokaryotic and eukaryotic communities exhibiting overlapping biogeographic patterns. Further, Central Park soils harboured nearly as many distinct soil microbial phylotypes and types of soil communities as we found in biomes across the globe (including arctic, tropical and desert soils). This integrated cross-domain investigation highlights that the amount and patterning of novel and uncharacterized diversity at a single urban location matches that observed across natural ecosystems spanning multiple biomes and continents.

1. Introduction

Soil is one of the most species-rich and diverse environments on the Earth, including members from all three domains of life (Bacteria, Archaea and Eukarya) [1,2]. Soil organisms can range in size from micrometres to centimetres and represent an amazing breadth of ecological strategies and metabolic capabilities [3,4], including some unique metabolisms that have been discovered only recently (e.g. [5]). These organisms have critical roles in terrestrial ecosystems and are responsible for a myriad of essential below-ground processes [6]. Nevertheless, the ecological attributes of many soil taxa (even very abundant bacterial and fungal taxa [7]) remain poorly understood, and the full-extent of the biological diversity found in soil remains unknown [8]. An improved understanding of the diversity that exists below-ground can help elucidate the ecological mechanisms underlying community structure and life-history traits of undescribed taxa. In addition, such below-ground assessments are needed to further develop conceptual models of the factors controlling microbial diversity and distribution patterns, as microbial ecology still lags behind plant and animal ecology in our ability to understand and predict biogeographic patterns [9]. We do know that patterns in microbial biogeography often differ from those observed for plants and animals due to
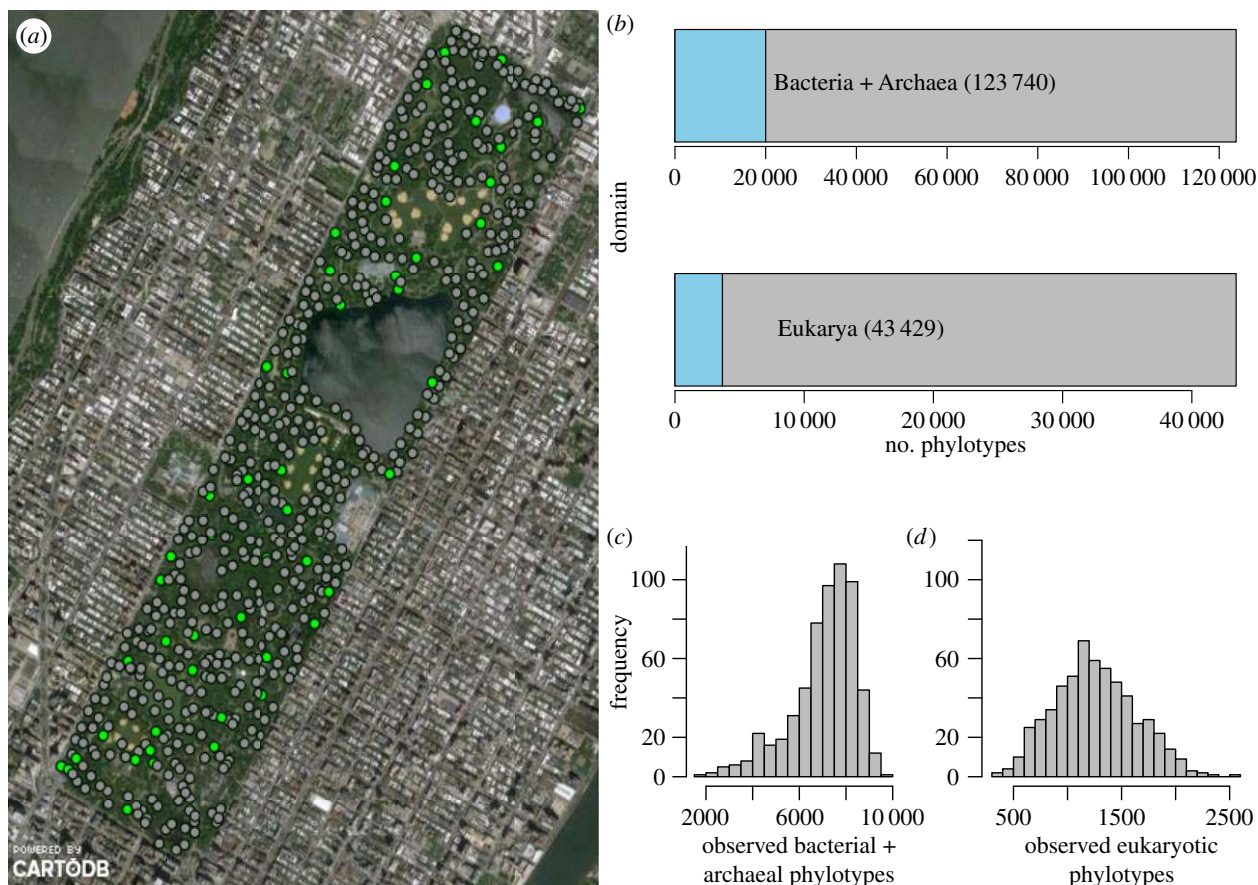
Figure 1. Central Park soils harbour high levels of undescribed biodiversity, with predictable biogeographic patterns. (a) Map of the 596 sampling locations in Central Park. Green coloured points indicate samples used in the comparison against the global sample set. (b) Of all bacterial and archaeal species, only 16.2% of sequences matched the Greengenes database. Likewise, of the eukaryotic species found in Central Park, only 8.5% of sequences matched the SILVA database. Histograms of (c) bacterial and archaeal and (d) eukaryotic observed number of phylotypes by samples (α-diversity) across the Park.

reasons of scale, phylogenetic breadth, taxonomic classification methods and dispersal capabilities [10], but a paucity of comprehensive assessments of the diversity patterns exhibited by microbes, particularly below-ground microbes, means that key knowledge gaps persist.

Recent advances in DNA sequencing methods have provided unprecedented insight into the biological diversity and the distribution patterns exhibited by soil taxa across multiple scales ranging from individual soil aggregates [11] to whole continents [12,13]. However, such studies have largely focused on individual groups (e.g. only bacteria), with far less attention paid to unicellular and multicellular eukaryotes despite increasing evidence that the diversity of soil fungi, protists and metazoa is likely far higher than often considered [14–16]. Indeed, there are few cross-domain assessments of below-ground diversity (but see [17,18]). Additionally, the vast majority of soil diversity studies have been conducted in natural settings, yet anthropogenic pressures now structure many ecosystems and it remains unclear whether there are consistent processes structuring soil biodiversity and biogeography in natural and urban ecosystems. In order to build on our understanding of soil diversity and biogeography patterns, we analysed soil samples collected from throughout Central Park in New York City, a highly managed, urban system.

Central Park is the most visited park, in the largest city in the USA [19] and an iconic site familiar to people worldwide. It is also ideally suited for investigations of below-ground diversity, because the soils found throughout the Park are highly variable in their habitat characteristics in part due to

the intensive development of the Park since its establishment in the mid-1800s [19]. The broad range of cover types and management practices (e.g. fertilizer and compost applications, mulching, irrigation) within the Park allowed us to examine the factors structuring soil communities across environmental gradients while holding climatic conditions nearly constant. Moreover, the Park's relatively small size (3.41 km²), allowed us to effectively sample the entire area (596 samples collected in total, with approx. one sample taken every 50 m of park land traversed in a regular grid), yielding a comprehensive 'snapshot' of cross-domain biodiversity and biogeographic patterns (figure 1a). Above-ground, Central Park harbours approximately 393 plant species, more than 250 species of vertebrates and more than 100 species of invertebrates [19]. There is no comparable estimate of the biodiversity found within the Park's soils; it is a *terra incognita* in one of the most frequently visited urban parks in the world.

Here, we investigate patterns of diversity and biogeography of Archaea, Bacteria, fungi, Protozoa, invertebrates and other eukaryotes across nearly 600 soil samples collected from Central Park in New York City, and then compare these patterns to a global soil biodiversity dataset. We quantified the diversity living within the soils via high-throughput sequencing of a hypervariable region of the 16S small subunit rRNA gene for the bacterial and archaeal analyses [20] and a comparable region of the 18S rRNA gene for the eukaryotic analyses. For reasons of consistency and for lack of a better definition that applies across all three domains [14,21], we define a phylotype as those taxa that share greater than or

equal to 97% sequence similarity in the targeted rRNA gene regions, following convention [22]. Using this definition, phylotypes could be considered equivalent to species, but we refer to them here as phylotypes to avoid confusion as we recognize that there are numerous definitions of what constitutes a species. This definition was applied across all three domains and yields a conservative estimate of species-level diversity compared to plant and animal surveys based on more traditional species delineations [23]. All samples were compared at an equivalent sequencing depth of 40 000 (16S or 18S rRNA gene) sequences per sample (nearly 50 million sequence reads or 7 500 000 000 nucleotides of data in total) to provide the most comprehensive assessment of soil diversity conducted to date. This study shows that much of the biodiversity below-ground remains undescribed, both within Central Park and from soils collected across global biomes; that the diversity in Central Park soils is comparable to diversity in soils collected from 'natural' ecosystems; and that below-ground biogeographic patterns are better predicted by the soil environment, rather than climate or geographical distance-factors that have traditionally been associated with plant and animal distribution patterns.

2. Material and methods

(a) Site description

Soils were collected from 596 locations across Central Park, New York City, USA on a single day. Central Park was established in 1857 and is 3.41 km^2, 0.80 km wide by 4.02 km long. The Park is not continuous and has a number of obstacles that we did not sample, including bodies of water, various buildings, a zoo and sports fields. The landscape is heterogeneous, ranging from large lawns to dense forests and a wide range of management regimes are employed throughout the Park. For this study, cover was classified as lawn (54%), tree (13%), herbaceous (18%), shrub (5%), other (10%) ('tree' sites had more than 10 cm diameter trees within a 10 m radius of the sample, and 'other' included mulch, bark, path and no vegetation).

(b) Sample collection and soil measurements

Samples were collected from approximately 50 transects running northwest to southeast across the Park. For each transect, samples were collected approximately every 50 m resulting in 10–15 samples per transect. At the time of sample collection, latitude and longitude, cover type and number of trees over 10 cm diameter and within a 10 m radius of the sampling site were recorded. Sample location was visualized using CartoDB (figure 1a) (cartodb.com). At each site location, four cores, each 2.54 cm diameter by 5 cm deep, were bulked to equal one sample per site. Soils were then sieved to 2 mm and carefully homogenized within 30 h of collection.

For each soil sample, pH, soil moisture, soil carbon and nitrogen concentrations and microbial biomass were determined on fresh soil. To measure soil pH, water and field-moist soil were mixed in a 1 : 1 volumetric ratio, allowed to stand for 10 min, and then pH was estimated in the supernatant using a benchtop pH meter. Gravimetric moisture (% water) of fresh soil was determined by oven drying to constant mass at 105°C. Total soil C and N content was determined on an elemental analyser (LECO, St Joseph, MI, USA). Microbial biomass was determined using the substrate-induced respiration (SIR) method. Briefly, 4 g dry weight equivalent soil per tube was incubated overnight at 20°C, before addition of 4 ml yeast solution (12 g yeast to 1 l H$_2$O). Soils were then incubated uncapped for 1 h, capped and flushed with CO$_2$-free air, and then finally incubated at 20°C for 5 h. Net CO$_2$ accumulation was measured on an infrared gas analyser. We report SIR biomass as the maximum CO$_2$ production rates (soil + substrate-derived); no conversion factors were used.

(c) Central Park community-level sequence analysis

To determine the diversity and composition of the soil community, genomic DNA was extracted from each soil sample using the MoBio 96-well extraction method [24]. Briefly, sterile cotton swabs were used to add each homogenized soil sample to the PowerSoil Bead Plate, and DNA was extracted following the instruction of the manufacturer with the modifications described previously [13]. DNA was amplified in triplicate using primers specific to either the 16S or 18S rRNA gene. A portion of the 16S rRNA gene was amplified using the Archaea- and Bacteria-specific primer set 515f/806r [20]. This 16S primer set is designed to amplify the V4–V5 region of both Archaea and Bacteria, has few biases against specific taxa and accurately represents phylogenetic and taxonomic assignment of sequences [25]. The 18S rRNA gene was amplified using the eukaryotic-specific primer set F1391 (5′-GTACACCGCCCGTC-3′) and REukBr (5′-TGATCCTTCTGCAGGTTCACCTAC-3′). The 18S primer set is designed to amplify the V9 hypervariable region of eukaryotes, with a focus on microbial eukaryotic lineages [26]. Amplicons were sequenced on two lanes of a 2 × 151 bp sequencing run on the Illumina HiSeq 2500 operating in Rapid Run Mode, following [3,27]. The raw forward read sequence data were demultiplexed and formatted for processing [28] using an in-house Python script. UPARSE was used for sequence clustering because it provides a relatively conservative estimate of microbial phylotype richness by reducing the number of spurious sequence clusters (i.e. operational taxonomic units or phylotypes), when compared with other commonly used sequence-processing pipelines [28]. In order to increase the computational efficiency of sequence processing, eukaryotic sequences were randomly subsampled to a common sequencing depth of 120 000 sequences per sample prior to running the UPARSE pipeline. Quality filtering was conducted by truncating sequences to 150 bp and using a maxee value of 0.5 (signifying that on average one nucleotide in every two sequences is incorrect). Filtered reads were dereplicated and unique sequences (i.e. singletons) were removed. These sequences were clustered into phylotypes following the UPARSE pipeline, which incorporates chimera checking into this step, and representative sequences for each phylotype were provided. Next, the raw demultiplexed sequences (78 141 936 16S rRNA and 70 402 319 18S rRNA gene sequences) were mapped to these representative sequences at the greater than or equal to 97% identity threshold, and 90% of 16S and 89% of 18S rRNA genes were successfully mapped to a phylotype. We recognize that the detected phylotypes are not necessarily active or living, and may represent inviable propagules, or may be derived from fragments of extracellular DNA. Prokaryotic phylotypes were classified to corresponding taxonomy using the RDP classifier [29] with a confidence threshold of 0.5, and eukaryotic phylotypes were classified to corresponding taxonomy using the top BLAST hit [30] as implemented in QIIME v. 1.6.0 [31], using default settings. During the 18S rRNA sequence processing, phylotypes classified only to the domain level or those without a BLAST hit were removed from downstream analyses. When conducting taxonomy assignments, the Greengenes 13_5 and SILVA 111 databases were used for prokaryotes and eukaryotes, respectively [32,33]. All samples were rarified to 40 000 randomly selected reads per sample, after samples were removed due to sampling error or falling below the rarified threshold, and 594 and 581 samples were included in downstream analyses of the prokaryotic and eukaryotic communities, respectively.

To calculate the proportion of phylotypes from Central Park that were represented in existing databases, we compared representative sequences from each phylotype against either the Greengenes or SILVA databases at the greater than or equal to 97% similarity threshold using USEARCH v. 7.0 [34]. Those sequences that successfully clustered with database sequences were considered to be representative of taxa archived in the respective databases. α-Diversity was determined from the number of phylotypes per sample or collection of samples (phylotype richness). To determine differences in taxonomic community composition across the Park, QIIME [31] was used to estimate pairwise dissimilarity between samples by calculating Bray–Curtis distances.

(d) Comparing Central Park soil diversity to global soil diversity

To compare the soil communities of Central Park to those communities found in other soils, we selected a 'global soil' sample set. Briefly, 52 soils representing a range of biomes from Alaska to Antarctica were selected from two previous studies [3,13]. The global soil sample set was compared to a randomly selected subset of Central Park soils (52 samples; figure 1a; electronic supplementary material, table S5). To characterize the bacterial and archaeal community sequences from the global soils and Central Park sample sets, raw sequences from both datasets were processed together. Briefly, 16S rRNA gene sequence reads were truncated to a common 90 bp, processed using methods described above and rarified to 40 000 sequences per sample. To characterize the eukaryotic communities of the global soils sample set, sequence data were obtained using the protocol described above on archived frozen samples, and sequenced on an Illumina MiSeq at the University of Colorado. Raw 18S rRNA gene sequences from both datasets were processed together using methods described above and rarified to 40 000 sequences per sample.

We used a number of metrics to compare soil biodiversity between the Central Park and global soil sample sets. The relative abundance of potentially pathogenic bacteria was calculated following Kembel *et al*. [35]. USEARCH [34] was conducted comparing each phylotype with a reference database of bacterial strains that are known human pathogens [36]. A phylotype was classified as a potential human pathogen if it shared greater than or equal to 97% sequence identity with a bacterial strain in the reference database (see [35] and references therein).

To compare the phylogenetic diversity between Central Park and the global soils, we selected phylotypes that appeared greater than 200 times in either of the sample sets and were classified to at least phylum-level taxonomy (hereon referred to as the most abundant microbial phylotype). This threshold selection constrains our analysis to phylotypes that represent greater than approximately 0.01% of all sequences and only those phylotypes that are relatively closely related to known microbial taxa, making our analyses more conservative by excluding rare phylotypes and minimizing the potential effects of PCR or sequencing errors. Remaining bacterial and archaeal phylotypes totalled 2497 and remaining eukaryotic phylotypes totalled 2342. Phylogenetic trees were built in order to assess whether specific lineages were uniquely represented in the Central Park or global datasets. To build the prokaryotic tree, the 2497 filtered sequences were first clustered to the Greengenes database at greater than or equal to 97% similarity in order to extract longer sequences and provide a more robust phylogeny. Sequences that matched the database (87% of sequences) were replaced with the longer Greengenes representative sequence. Those representative sequences and the remaining original sequences that did not match were used to build a phylogenetic tree. Sequences were aligned using PyNAST [37] and highly conserved regions were filtered in QIIME, as they are unhelpful in building the phylogeny. The maximum-likelihood tree was computed using Fasttree [38]. The archaeal sequences were used to root the tree. To build the eukaryotic tree, the original 150 bp sequences were used to build a phylogenetic tree. Sequences were aligned, filtered and used to compute a tree as above. Deeply divergent nematode sequences were used to root the tree. Both trees were coloured and formatted using GraPhlAn; colour was added to highlight phylotypes shared between datasets and phylotypes only found in either Central Park or the global soil sample sets. The trees are not intended to represent the detailed evolutionary history among phylotypes, but rather they highlight the phylogenetic diversity shared between Central Park and the global soils.

(e) Statistical analyses

To assess the relationship between community similarity and environmental conditions, we used two different metrics to quantify community similarity, Bray–Curtis distances (which are based on relative abundances) and Jaccard distance (a presence–absence metric). However, since the results were nearly identical with the two metrics (electronic supplementary material, table S3), our discussion focuses on results from the analyses of Bray–Curtis distances. The pairwise distances in community similarity from the prokaryotic and eukaryotic communities were compared to each other and to edaphic characteristics using Mantel tests based on Spearman's rank correlations. Partial Mantel tests were used to test for relationships between any two distance matrices while controlling for a third. Likewise, multiple regressions on the pairwise Bray–Curtis distances were used to further explore relationships between three variables. Co-occurrence patterns between the prokaryotic and eukaryotic communities were tested using Spearman's rank correlations between OTUs that occurred in at least 25% of the samples and had a $\rho > 0.6$ and p-value of less than 0.001 (adjusted using the FDR method) [39]. Differences in the proportion of potential pathogens were tested using Mann–Whitney tests, with an FDR correction. All analyses were performed using the R program v. 3.0.0 using Vegan and Ecodist packages.

3. Results and discussion

(a) Central Park soil diversity is novel and diverse

A total of 122 081 bacterial, 1659 archaeal and 43 429 eukaryotic phylotypes (figure 1b) were found across Central Park, with Archaea representing a relatively small proportion of the 'prokaryotic' community (approx. 1.4% of the 16S rRNA gene sequences). Not only did Central Park harbour high levels of below-ground diversity, we found that most of these phylotypes were undescribed. Briefly, we compared the sequences of the prokaryotic and eukaryotic phylotypes from Central Park to sequences deposited in the most comprehensive databases available to date and found that only 8.5–16.2% of the Central Park phylotypes had matches to their respective databases (at the greater than or equal to 97% similarity level; figure 1b). Most strikingly, of the phylotypes that matched their respective reference databases, many are known only by their rRNA gene sequences and the taxonomic assignments are restricted to only the phylum level of resolution (but we acknowledge that more phylotypes may be known and previously described via morphology but still lack representation in the molecular databases, an issue particularly relevant for eukaryotic microbes [40]). Of the dominant phylotypes found within the Central Park soils all were relatively underrepresented in the databases

(electronic supplementary material, table S1), and a few phyla (i.e. *Proteobacteria*, *Verrucomicrobia*, Rhizara and Stramenopiles) represented a particularly large fraction of the novel diversity and would be good targets for future survey efforts. Together these results support previous speculation that most soil diversity remains largely unexplored [41], meaning the reference databases are far from complete and require studies such as ours to minimize their over-representation of well-studied and cultured lineages.

Not only did Central Park harbour large numbers of pro-karyotic and eukaryotic phylotypes (high γ-diversity, *sensu* [42]) but also harbour a broad range of below-ground community types (high β-diversity). Individual soil samples shared relatively few phylotypes in common and most phylotypes were restricted in their distribution across the Park. This is evident from a comparison of the estimated γ-diversity (figure 1*b*) to the mean α-diversity per sample (7041 bacterial and archaeal phylotypes, and 1257 eukaryotic phylotypes per sample; figure 1*c*,*d*). Because α-diversity is so much lower than γ-diversity, there is clearly a high degree of variability in community composition from sample to sample, with any randomly selected pair of samples sharing on average only 19.3% of their bacterial and archaeal phylotypes, and 13.5% of their eukaryotic phylotypes. This high degree of variability in community composition was evident even when we compared the relative abundances of major taxonomic groups across the collected soils. For example, the relative abundances of the dominant bacterial phyla, including *Proteobacteria*, *Acidobacteria*, *Verrucomicrobia*, *Actinobacteria*, *Bacteroidetes*, varied by as much as 38-fold between samples, while the relative abundance of Archaea ranged from 0 to 13% across the Park (electronic supplementary material, table S2). Likewise, eukary-otic communities were dominated by Rhizara, *Apicomplexa*, stramenopiles, fungi and various metazoan taxa, whose relative abundance varied as much as 36-fold between samples (electronic supplementary material, table S3). Clearly, below-ground community composition was highly variable across the Park yet, in contrast to the patterns commonly observed for plant and animal communities [43], geographical distance was not a significant predictor of below-ground community structure ($r_M = 0.06$ and $r_M = 0.03$ for prokaryotic and eukary-otic communities, respectively) (electronic supplementary material, table S3 and figure S2*a*,*b*). In other words, sites closer together did not harbour communities more similar in composition than sites located further apart. This disconnect between geographical and community distance is probably due to the mosaic and discontinuous nature of the cover types and management regimes across the Park where soil con-ditions can change abruptly at our 50 m sampling resolution (electronic supplementary material, figure S1*a*−*f* and table S5).

Despite the high degree of heterogeneity in bacterial and archaeal community types across the Park, the biogeographic patterns exhibited by these soil communities were predictable from environmental characteristics. The best environmental pre-dictor of the patterns in prokaryotic community composition was soil pH ($r_M = 0.45$; $p < 0.001$) (electronic supplementary material, figure S2*c*). Similar patterns have been reported pre-viously from non-urban ecosystems, where pH has been shown to be an important driver of soil bacterial community composition [13,44]. The strong influence of soil pH on bacte-rial biogeography was particularly important in driving the proportional abundance of *Acidobacteria* across the Park (elec-tronic supplementary material, figure S2*e*). No other measured

environmental variable was significantly correlated with the prokaryotic community distribution patterns (all $r_M < 0.1$) (electronic supplementary material, table S3), including plant cover type which suggests that, within Central Park, plant com-munity composition was not a good predictor of below-ground diversity patterns. While other unmeasured variables may also contribute to the observed biogeographic patterns, our results emphasize the importance of soil pH as a determinant of pro-karyotic biodiversity across soils at local as well as regional to global scales. When diversity patterns were assessed via the Jaccard index, we observed similar patterns (electronic sup-plementary material, table S3), suggesting that the observed biogeographic patterns are not only associated with differences in the relative abundance of these groups but also the presence or the absence of phylotypes in these communities.

Eukaryotic community composition was not as strongly pre-dicted from soil characteristics nor was it significantly correlated with plant cover type; soil pH was the only measured variable that significantly correlated with eukaryotic biogeographic patterns across the Park, but it was not a particularly strong pre-dictor ($r_M = 0.20$; $p < 0.001$; electronic supplementary material, figure S2*d*). Instead, the best predictor of eukaryotic community patterns was prokaryotic community composition ($r_M = 0.53$, $p = 0.001$), regardless of the distance metric employed (electro-nic supplementary material, figure S2*f*). This correspondence between the prokaryotic and eukaryotic communities was even stronger when we controlled for the effect of soil pH (electronic supplementary material, table S3), highlighting that this relationship is not solely a product of shared environ-mental preferences. There are numerous direct or indirect associations between prokaryotic and eukaryotic taxa that could yield the shared spatial distribution patterns observed here. For example, when we examined co-occurrence patterns between individual phylotypes we found that the relative abun-dances of a number of bacterial phylotypes (namely members of the *Acidobacteria*, *Gammaproteobacteria* and *Verrucomicrobia* phyla) exhibited strong positive correlations with the relative abundances of individual eukaryotic phylotypes (particularly various rhizarian and fungal taxa; electronic supplementary material, table S4). These associations could be a product of trophic interactions between predatory protists and their bacterial prey [45], a product of direct symbioses (e.g. the relationships between fungi and bacteria [46]), or simply shared environmental drivers. The shared biogeographic patterns of prokaryotic and eukaryotic communities demon-strates that there are numerous direct and indirect associations between soil organisms, and unravelling these relationships will be critical to building a more integrated understanding of below-ground ecology.

(b) Central Park soil biodiversity is similar to global soil biodiversity

Given that Central Park was found to harbour large numbers of bacterial, archaeal and eukaryotic taxa and a broad range of community types, we asked how the below-ground diver-sity found in Central Park compares to soils collected from a wide range of soil types, climate zones and biome types. Using 52 randomly selected Central Park soil samples and 52 'global' soil samples (collected from biomes as distinct as Antarctic cold deserts, tropical forests, temperate forests, arctic tundra and grasslands; figure 1*a*; electronic supplemen-tary material, figure S3), we assessed and compared the
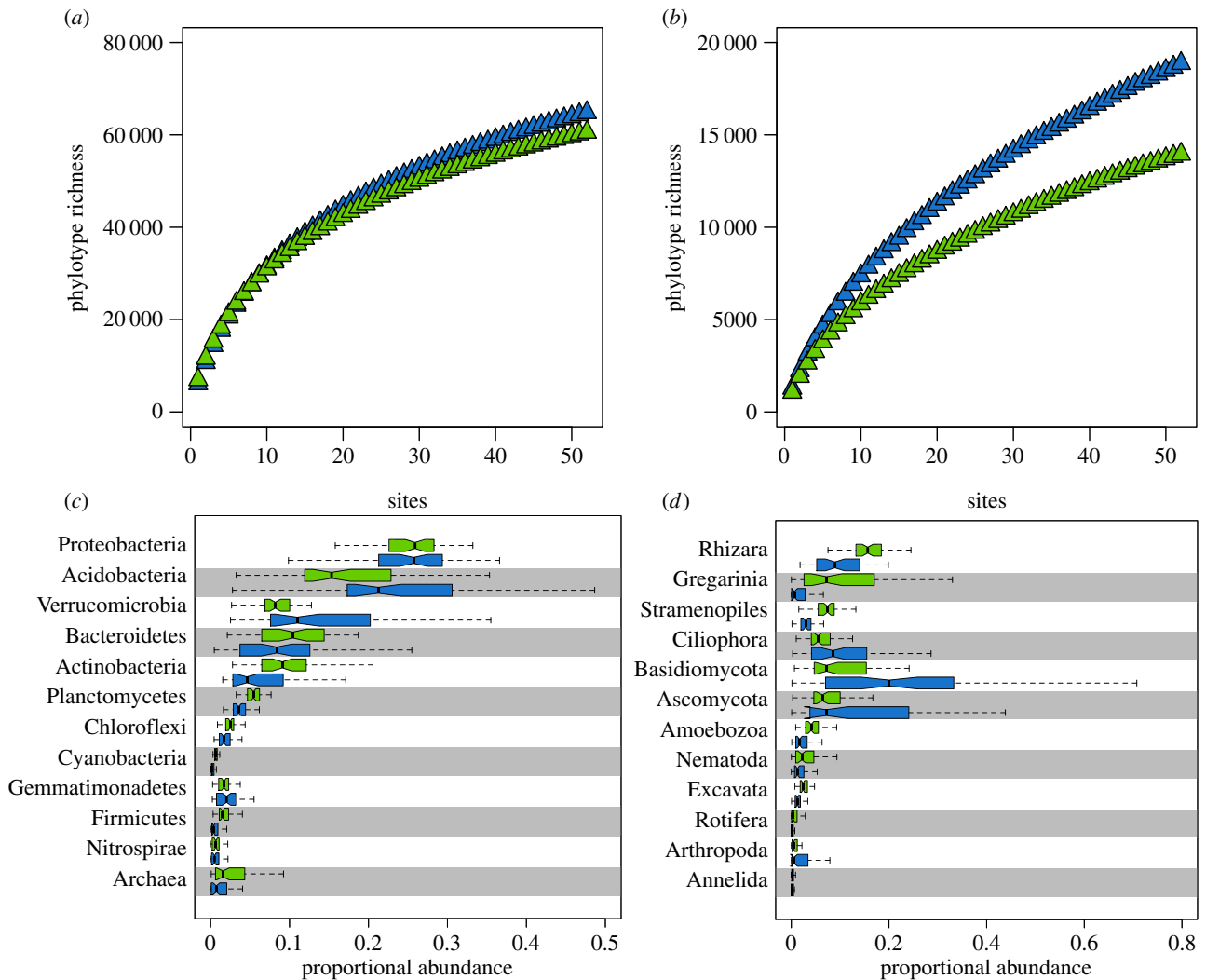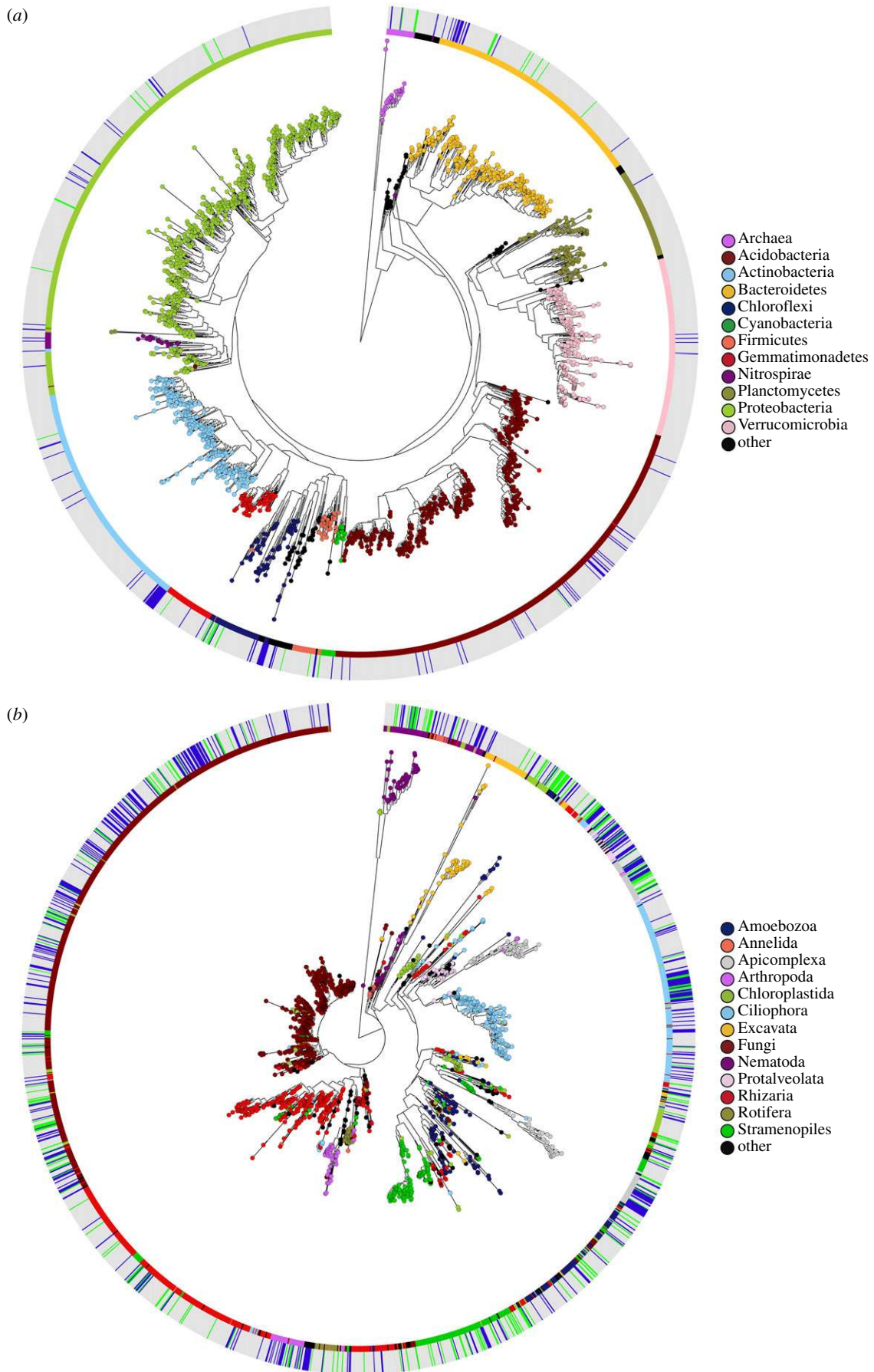
Figure 2. Central Park soil diversity is similar to soil communities from other biomes. Phylotype accumulation curves for (a) bacterial and archaeal communities, and (b) eukaryotic communities from Central Park (green) and global soils (blue). Relative abundances of the most dominant (c) bacterial and archaeal taxa, and (d) eukaryotic taxa from Central Park (green) and global soil sample set (blue).

diversity of all 104 samples using identical methods. Regardless of the diversity metric employed, Central Park soil diversity was markedly similar to soil communities from other biomes. First, the below-ground diversity found in Central Park was similar in magnitude to the diversity observed in the global dataset, with the Park having only 6.5% fewer prokaryotic phylotypes and 26% fewer eukaryotic phylotypes than what was observed in the global sample set (figure 2a,b). Likewise, when we compared the relative abundances of the dominant bacterial, archaeal and eukaryotic phylotypes we found a surprising amount of overlap between the Central Park soils and the global sample set (figure 2c,d). This high degree of overlap between the Central Park below-ground communities and the communities represented by our global sample set is likely a product of Central Park having such a broad range in soil edaphic characteristics. In fact, the range in measured soil characteristics (including pH, carbon and nitrogen concentrations) nearly matched the range observed across the global sample set (electronic supplementary material, table S5). The management practices used in Central Park not only promote a diverse array of soil conditions, but they also appear to result in those urban soils harbouring a breadth of below-ground taxa and community-type, rivalling that found collectively in soils from across the globe. Moreover, these results suggest that,

unlike plant and animal communities, climate is not a dominant driver of soil biogeography given that conditions within Central Park span a very narrow portion of the climatic gradient represented by our global sample set which includes sites from Antarctic cold deserts to tropical Peru.

If we compare the overlap between the Central Park soils and the soils collected from across the globe at finer scales of taxonomic resolution, a similar pattern emerges. Of the most abundant bacterial and archaeal phylotypes from the two sample sets (those that represent greater than approx. 0.01% of sequences, see Material and methods) 94.7% of the 2497 phylotypes were shared between Central Park and the global soils, 1.3% were found only in Central Park, and 4.0% were found only in the global soils (figure 3a). Most of the bacterial phylotypes not found in Central Park were those phylotypes common in the extremely high pH soils found in the desert soils (both cold and hot deserts; electronic supplementary material, figure S4). Those phylotypes not found in the global sample set were from a variety of genera (e.g. *Candidatus*, Nitrososphaera, *Sphingobacterium* sp. and *Rhodospirillaceae*) that are probably associated with the compost added to the Park's soils [47]. Of the most abundant eukaryotic phylotypes from the two sample sets 73% of the 2342 phylotypes were shared between Central Park and the global soils, 9% were found only in Central Park and

7

rspb.royalsocietypublishing.org   Proc. R. Soc. B 281: 20141988

(a)

(b)

Archaea
Acidobacteria
Actinobacteria
Bacteroidetes
Chloroflexi
Cyanobacteria
Firmicutes
Gemmatimonadetes
Nitrospirae
Planctomycetes
Proteobacteria
Verrucomicrobia
other

Amoebozoa
Annelida
Apicomplexa
Arthropoda
Chloroplastida
Ciliophora
Excavata
Fungi
Nematoda
Protalveolata
Rhizaria
Rotifera
Stramenopiles
other

Figure 3. Phylogenetic overlap between Central Park and global soil communities. The phylogenetic trees of (a) bacterial and archaeal phylotypes, and (b) eukaryotic phylotypes show approximate evolutionary relationships among taxa observed in Central Park and the global soil datasets. Dominant clades are indicated with unique colours (inner ring). The outer ring is coloured grey to represent if a phylotype was found in both Central Park and in the global soil sample set, green if found only in Central Park or blue if found only in the global soil sample set.

18% were found only in global soils (figure 3b). Many of the eukaryotic phylotypes not found in Central Park were individual fungal phylotypes including mycorrhizal phyla whose biogeographic patterns are probably determined by associations with vegetation types (e.g. boreal forest) not found in Central Park [48].

Although we found a high degree of overlap in the composition of communities from Central Park and those from other biomes, there are clearly numerous taxonomic groups, which differ in abundance between Central Park and other biomes. However, one particular group that differed in relative abundance between the two sample sets were those phylotypes whose sequences were close matches to human pathogens [35,36]. While sequences matching potential human pathogens were relatively rare in all soils, human-associated potential pathogens were over two times more abundant in the Central Park soils ($p < 0.001$) (electronic supplementary material, figure S5a). In particular, *Staphylococcus saprophyticus*, *Salmonella enterica* and *Citrobacter koseri*, and a well-known spore-former *Bacillus anthracis* were all consistently more abundant in the Central Park soils (all $p < 0.001$) (electronic supplementary material, figure S5b). We want to stress that the presence of potential pathogen sequences does not indicate the presence of a disease-causing organism in the soil, rather this finding highlights a significant difference between soil bacterial communities found in more natural systems and those in Central Park. Furthermore, we do not know why the relative abundance of phylotypes related to known pathogens is higher in Central Park than in other soils; it could be a product of the large human populations surrounding Central Park and anthropogenic pressures or due to the history of intensive disturbance regimes within the Park [49]. Clearly, it is worth investigating whether this pattern is widespread and if soils in urban parks are more likely to harbour potential pathogens than soils from non-urban locations.

Our work highlights that most of the diversity found in soil remains undescribed, and although 'everything' is not likely to be 'everywhere' (*sensu* Baas-Becking, [50]), we can find nearly as many different soil species and community types within the 3.41 km^2 area of Central Park as we would find if we travelled around the world collecting a broad array of soil types. We do not have the data available to quantitatively compare the microbial biogeographic patterns to patterns in plant and animal communities across these sites and due to differences in how species are defined, and it would be difficult to directly compare biogeographic patterns even if we had such data [10]. Nevertheless, it is reasonable to assume that 95% of the plant or animal species found in Central Park do not also occur across the many other ecosystems we sampled from, including tropical rainforests, tundra and deserts. Here, we assess local- to global-scale distribution patterns of soil microbes, and the next steps should move beyond descriptions of the taxa found in soil, to understand the functions of these taxa in soil [1,2] and how they interact with one another [51]. By doing so, we will not only build on our basic understanding of these vital communities found below-ground, we may also learn how to actively manage these communities [52] to promote soil fertility, reduce soil pathogen loads and restore degraded lands.

References

1. Fierer N et al. 2012 Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl Acad. Sci. USA* **109**, 21 390–21 395. (doi:10.1073/pnas.1215210110)
2. Wall DH, Bardgett RD, Kelly E. 2010 Biodiversity in the dark. *Nat. Geosci.* **3**, 297–298. (doi:10.1038/ngeo860)
3. Dinsdale EA et al. 2008 Functional metagenomic profiling of nine biomes. *Nature* **452**, 629–632. (doi:10.1038/nature06810)
4. Torsvik V, Øvreås L. 2002 Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* **5**, 240–245. (doi:10.1016/S1369-5274(02)00324-7)
5. Humbert S, Tarnawski S, Fromin N, Mallet M-P, Aragno M, Zopfi J. 2010 Molecular detection of anammox bacteria in terrestrial ecosystems: distribution and diversity. *ISME J.* **4**, 450–454. (doi:10.1038/ismej.2009.125)
6. Wall DH, Bardgett RD, Behan-Pelletier V, Herrick JE, Jones H, Ritz K, Six J, Strong DR, van der Putten WH. 2012 *Soil ecology and ecosystem services*. USA: Oxford University Press (cited 5 November 2013).
7. 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. (doi:10.1126/science.1243768)
8. Decaëns T. 2010 Macroecological patterns in soil communities. *Glob. Ecol. Biogeogr.* **19**, 287–302. (doi:10.1111/j.1466-8238.2009.00517.x)
9. Martiny JB et al. 2006 Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112. (doi:10.1038/nrmicro1341)
10. Fierer N, Lennon J. 2001 The generation and maintenance of diversity in microbial communities. *Am. J. Bot.* **98**, 439–448. (doi:10.3732/ajb.1000498)
11. Davinic M, Fultz LM, Acosta-Martinez V, Calderón FJ, Cox SB, Dowd SE, Allen VG, Zak JC, Moore-Kucera J. 2012 Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. *Soil Biol. Biochem.* **46**, 63–72. (doi:10.1016/j.soilbio.2011.11.012)
12. Wu T, Ayres E, Bardgett RD, Wall DH, Garey JR. 2011 Molecular study of worldwide distribution and diversity of soil animals. *Proc. Natl Acad. Sci. USA* **108**, 17 720–17 725. (doi:10.1073/pnas.1103824108)
13. Lauber CL, Hamady M, Knight R, Fierer N. 2009 Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120. (doi:10.1128/AEM.00335-09)
14. Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. 2012 Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* **27**, 233–243. (doi:10.1016/j.tree.2011.11.010)
15. Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N. 2013 Global biogeography of highly diverse protistan communities in soil. *ISME J.* **7**, 652–659. (doi:10.1038/ismej.2012.147)
16. Robeson MS, King AJ, Freeman KR, Birky CW, Martin AP, Schmidt SK. 2011 Soil rotifer communities are extremely diverse globally but spatially

autocorrelated locally. *Proc. Natl Acad. Sci. USA* **108**, 4406–4410. (doi:10.1073/pnas.1012678108)

17. Fierer N *et al*. 2007 Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of Bacteria, Archaea, fungi, and viruses in soil. *Appl. Environ. Microbiol.* **73**, 7059–7066. (doi:10.1128/AEM.00358-07)

18. Delmont TO *et al*. 2012 Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J.* **6**, 1677–1687. (doi:10.1038/ismej.2011.197)

19. About the Central Park Conservancy. 2010 See http://www.centralparknyc.org/about.

20. 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. (doi:10.1038/ismej.2010.171)

21. Konstantinidis KT, Ramette A, Tiedje JM. 2006 The bacterial species definition in the genomic era. *Phil. Trans. R. Soc. B* **361**, 1929–1940. (doi:10.1098/rstb.2006.1920)

22. Stackebrandt E, Goebel BM. 1994 Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *IJSEM.* **44**, 846–849. (doi:10.1099/00207713-44-4-846)

23. Mallet J. 1995 A species definition for the modern synthesis. *Trends Ecol. Evol.* **10**, 294–299. (doi:10.1016/0169-5347(95)90031-4)

24. Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R. 2012 Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J.* **6**, 1007–1017. (doi:10.1038/ismej.2011.159)

25. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. 2007 Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* **35**, e120. (doi:10.1093/nar/gkm541)

26. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. 2009 A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **4**, e6372. (doi:10.1371/journal.pone.0006372)

27. Caporaso JG *et al*. 2012 Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624. (doi:10.1038/ismej.2012.8)

28. Edgar RC. 2013 UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998. (doi:10.1038/nmeth.2604)

29. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007 Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267. (doi:10.1128/AEM.00062-07)

30. Altschul S. 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402. (doi:10.1093/nar/25.17.3389)

31. Caporaso JG *et al*. 2010 QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336. (doi:10.1038/nmeth.f.303)

32. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013 The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. (doi:10.1093/nar/gks1219)

33. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012 An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of Bacteria and Archaea. *ISME J.* **6**, 610–618. (doi:10.1038/ismej.2011.139)

34. Edgar RC. 2010 Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461. (doi:10.1093/bioinformatics/btq461)

35. Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM, Bohannan BJ, Brown GZ, Green JL. 2012 Architectural design influences the diversity and structure of the built environment microbiome. *ISME J.* **6**, 1469–1479. (doi:10.1038/ismej.2011.211)

36. UK Health and Safety Executive Advisory Committee on Dangerous Pathogens. 2004 *The approved list of biological agents*. UK Health and Safety Executive. (http://www.hse.gov.uk/pubns/misc208.pdf)

37. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010 PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267. (doi:10.1093/bioinformatics/btp636)

38. Price MN, Dehal PS, Arkin AP. 2009 FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650. (doi:10.1093/molbev/msp077)

39. Verhoeven KJF, Simonsen KL, McIntyre LM. 2005 Implementing the false discovery rate control: increasing your power. *Oikos* **108**, 643–647. (doi:10.1111/j.0030-1299.2005.13727.x)

40. Creer S *et al*. 2010 Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Mol. Ecol.* **19**(Suppl. 1), 4–20. (doi:10.1111/j.1365-294X.2009.04473.x)

41. Pace NR. 1997 A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740. (doi:10.1126/science.276.5313.734)

42. Whittaker RH. 1960 Vegetation of the Siskiyou Mountains, Oregon and California. *Eco. Mono.* **30**, 279–338. (doi:10.2307/1943563)

43. Gaston KJ. 2000 Global patterns in biodiversity. *Nature* **405**, 220–227. (doi:10.1038/35012228)

44. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010 Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340–1351. (doi:10.1038/ismej.2010.58)

45. Adl M, Gupta V. 2006 Protists in soil ecology and forest nutrient cycling. *Can. J. For. Res.* **36**, 1805–1817. (doi:10.1139/x06-056)

46. Frey-Klett P, Garbaye J, Tarkka M. 2007 The mycorrhiza helper bacteria revisited. *New Phytol.* **176**, 22–36. (doi:10.1111/j.1469-8137.2007.02191.x)

47. Neher D, Weicht TR, Bates SB, Leff JW, Fierer N. 2013 Changes in bacterial and fungal communities across compost recipes, preparation methods, and composting times. *PLoS ONE* **8**, e79512. (doi:10.1371/journal.pone.0079512)

48. Kivlin SN, Hawkes CV, Treseder KK. 2011 Global diversity and distribution of arbuscular mycorrhizal fungi. *Soil Biol. Biochem.* **43**, 2294–2303. (doi:10.1016/j.soilbio.2011.07.012)

49. Abawi G, Widmer T. 2000 Impact of soil health management practices on soilborne pathogens, nematodes and root diseases of vegetable crops. *Appl. Soil Ecol.* **15**, 37–47. (doi:10.1016/S0929-1393(00)00070-6)

50. O'Malley MA. 2007 The nineteenth century roots of 'everything is everywhere'. *Nat. Rev. Microbiol.* **5**, 647–651. (doi:10.1038/nrmicro1711)

51. Faust K, Raes J. 2012 Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538–550. (doi:10.1038/nrmicro2832)

52. De Vries FT *et al*. 2013 Soil food web properties explain ecosystem services across European land use systems. *Proc. Natl Acad. Sci. USA* **110**, 14 296–14 301. (doi:10.1073/pnas.1305198110)