

Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere

A. Jumpponen^{1,2} and K. L. Jones^{2,3}

¹Division of Biology, Kansas State University, Manhattan, KS 66506, USA; ²Ecological Genomics Institute, Kansas State University, Manhattan, KS 66506, USA; ³Environmental Health Sciences, University of Georgia, Athens, GA 30602, USA

Summary

Author for correspondence:

A. Jumpponen

Tel: +1 785 532 6751

Email: ari@ksu.edu

Received: 16 April 2009

Accepted: 19 June 2009

New Phytologist (2009) **184**: 438–448

doi: 10.1111/j.1469-8137.2009.02990.x

Key words: endophyte, epiphyte, foliar, fungi, *Quercus*, urbanization.

- This study targeted the fungal communities in the phyllosphere of *Quercus macrocarpa* and compared the fungal species richness, diversity and community composition among trees located within and outside a small urban center using recently developed 454 sequencing and DNA tagging.
- The results indicate that the fungal phyllosphere communities are extremely diverse and strongly dominated by ascomycetes, with *Microsphaeropsis* [two Operational Taxonomic Units (OTUs); 23.6%], *Alternaria* (six OTUs; 16.1%), *Epicoccum* (one OTU; 6.0%) and *Erysiphe* (two OTUs; 5.9%) as the most abundant genera.
- Although the sequencing effort averaged 1000 reads per tree and detected nearly 700 distinct molecular OTUs at 95% internal transcribed spacer 1 similarity, the richness of the hyperdiverse phyllosphere communities could not be reliably estimated as nearly one-half of the molecular OTUs were singletons.
- The fungal communities within and outside the urban center differed in richness and diversity, which were lower within the urban development. The two land-use types contained communities that were distinct and more than 10% of the molecular OTUs differed in their frequency.

Introduction

Little is known about microbial communities in urban environments. They are thought to retain low biodiversity (Blair, 1999, 2004; Cincotta & Engelman, 2000; McKinney, 2002), although some empirical studies indicate otherwise and show that urban areas can be organismally quite diverse (Kühn *et al.*, 2004; Wania *et al.*, 2006). Although humans consciously manipulate the macro-organisms in urban greeneries, control of the microbial communities is usually indirect (McKinney & Lockwood, 1999). This may apply particularly well to foliar microbes that are ubiquitous and diverse in many environments (for example, Lambais *et al.*, 2006; Arnold & Lutzoni, 2007), but often go unseen unless the host plants are diseased.

The phyllosphere – that is, the living leaf as a whole including the inter- and intracellular interior spaces as well as the surface (Carroll *et al.*, 1977) – provides a habitat for a variety of fungi that occupy leaves superficially (epiphytes) or intracellularly (endophytes) (Santamaria & Bayman, 2005).

Multiple studies have documented that temperate plants harbor numerous fungal associates (see reviews by Saikkonen *et al.*, 1998 and Rodrigues *et al.*, 2009). The foliar endophytes, in particular, have received substantial attention as a potentially hyperdiverse fungal community (Arnold *et al.*, 2000, Arnold, 2007; Arnold & Lutzoni, 2007). The epiphytic communities probably add to this diversity further, underlining the importance of foliage as a habitat for hyperdiverse communities (Santamaria & Bayman, 2005).

In this study, our primary goals were to test whether or not the fungal communities in the *Quercus macrocarpa* phyllosphere differ in their richness, diversity or composition among the urban and rural stands. Furthermore, we aimed to identify taxa that were responsive to urbanization by showing either increases or decreases in their abundances in the urban environments. To efficiently utilize high-throughput sequencing, we used the recently developed massively parallel 454 pyrosequencing (Margulies *et al.*, 2005), and combined it with DNA tagging (see, for example, Hamady *et al.*, 2008; Meyer *et al.*, 2008). By doing so, we were able to independently and

Table 1 Properties of the six stands sampled for *Quercus macrocarpa* phyllosphere fungal communities

	City Park	Ahearn (KSU)	Beach Museum (KSU)	Kings Creek	Håkanson Homestead	Shane Creek
Land use	Urban	Urban	Urban	Rural	Rural	Rural
Location	N39°10'58" W96°35'25"	N39°11'16" W96°34'42"	N39°11'46" W96°34'53"	N39°06'21" W96°36'01"	N39°06'21" W96°34'27"	N39°06'42" W96°33'49"
Elevation (m asl)	342	331	329	337	333	354
Stand structure	Street side row	Street side row	Single or clustered trees	Hedgerow	Single or clustered trees	Hedgerow
Ground vegetation	Lawn	None or lawn	None or lawn	None or mixed prairie	None or mixed prairie	Mixed prairie

KSU, Kansas State University.

expediently sequence and characterize fungal communities in 54 leaf samples representing 18 trees.

Materials and Methods

Study site descriptions

The city of Manhattan is located in north-eastern Kansas (USA) in the Flint Hills region. Manhattan has *c.* 50 000 residents and 22 000 students, representing a mobile, transient population. *Quercus macrocarpa* Michx (bur oak) is a common native tree that is occasionally used as an ornamental, but is relatively infrequent because of its slow growth. Bur oaks occur sufficiently frequently to allow for comparisons among land uses. We selected three planted and managed stands within the Manhattan city limits (urban) and three native stands without regular park management regimes in a rural environment (Table 1) located in an undisturbed prairie within the Konza Prairie Long-Term Ecological Research site (<http://www.konza.ksu.edu/>). To our knowledge, none of the selected trees had been treated specifically for pathogens or parasites, but the urban sites had been regularly fertilized as a result of lawn and park management. We believe that the city of Manhattan was established in what would have been considered as a continuous tallgrass prairie before European settlement. The Konza Prairie sites are located within a landscape matrix presently set aside as a nature preserve and owned by the Nature Conservancy. This area has not been subject to agriculture, but was used for low-intensity grazing before preserve establishment. The stand characteristics were matched as closely as possible among the two land-use types. As a result, our native stands represented hedgerows and clustered trees. A total of three trees was selected at each of three sites for a total of nine trees in each land-use type.

Sampling, DNA extraction, PCR amplification and sequencing

In July 2007, one random inner canopy shade leaf was collected from three distinct 60° sectors within each of the three focal trees at the three sites at each of the two land-use types. Although genomic DNA was extracted and PCR amplicons were produced separately for each leaf, we considered a tree

as our experimental unit. This resulted in a total of 18 experimental units across the study ($n = 18$). The excised leaves were stored on ice in a plastic bag until transport to the laboratory for further processing. In the laboratory, the whole leaves were agitated in 0.1% Triton-X for 30 s and rinsed three times in deionized, sterile H₂O to remove nonadhering superficial fungal particles (spores and hyphal fragments). The leaves were blotted dry with clean paper towels and four disks per leaf were removed with a cork borer (diameter, 1 cm) for a total surface area of approximately 3.14 cm² per leaf. Areas of obvious and visible necrosis or insect damage were avoided in harvesting of the leaf disks. The excised disks from each leaf were pooled into a MoBio Bead Tube (UltraClean Soil DNA, MoBio Laboratories, Carlsbad, CA, USA) in the bead solution amended with two 2.4 mm zirconia beads for better homogenization. The leaf disks were homogenized in a FastPrep instrument (Qbiogene, Carlsbad, CA, USA) at maximum speed (setting 6.0) for 40 s. This effectively homogenized the disks to a point that no visually recognizable fragments remained. From that point on, total DNA was extracted as instructed by the manufacturer of the UltraClean Soil DNA kit and eluted in 100 µl of buffer S5. All templates were quantified with an ND1000 spectrometer (NanoDrop Technologies, Wilmington, DE, USA) and adjusted to a final concentration of 2.5 ng µl⁻¹. An initial optimization indicated that PCR amplicons were reliably produced by up to 10 ng of template in a 50 µl reaction. Accordingly, 10 ng (or 4 µl) of the template per reaction were used to generate the amplicons for sequencing.

For direct 454 sequencing of the fungal internal transcribed spacer 1 (ITS1) amplicons using massively parallel sequencing (MPS), we synthesized primer constructs that incorporated the MPS primers (Margulies *et al.*, 2005) and the fungus-specific (ITS1F; Gardes & Bruns, 1993) or universal (ITS2; White *et al.*, 1990) primers (Table S1, see Supporting Information). These primer constructs combined 454-sequencing primer (A-primer) and the reverse primer (ITS2) with a five base pair (bp) DNA tag for post-sequencing sample identification in between, or the DNA capture bead anneal primer (B-primer) for the emulsion PCR (emPCR) and the forward primer (ITS1F). This primer choice resulted in reverse sequence across the ITS1 region and minimized sequencing of the conserved 3'-end of the small subunit of the ribosomal

RNA gene. To allow simultaneous sequencing, optimally performing primers were selected through screening with an ascomycete (*Saccharomyces cerevisiae*), a basidiomycete (*Agaricus bisporus*) and an environmental sample (a soil DNA extract deemed amplifiable in other experiments) from the total of 96 synthesized primers.

After optimization for template concentration (above), $MgCl_2$ concentration and annealing temperature, fungal ITS1 was PCR amplified in 50 μ l reactions. Each sample was amplified in three separate reactions to account for heterogeneous amplification from the environmental template. Each reaction contained final concentrations or absolute amounts of reagents as follows: 200 nM of each of the forward and reverse primers, 10 ng of the template DNA, 200 μ M of each deoxynucleotide triphosphate, 2.5 mM $MgCl_2$, 1 unit of Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA) and 5 μ l of PCR buffer (Platinum Blue; Invitrogen). The PCR cycle parameters consisted of an initial denaturation at 94°C for 3 min, then 25 cycles of denaturation at 94°C for 1 min, annealing at 54°C for 1 min and extension at 72°C for 2 min, followed by a final extension step at 72°C for 10 min. All PCRs were performed in 96-well PCR plates on a Master-Cycler (Eppendorf, Hamburg, Germany). Possible PCR amplification of contaminants was determined using a blank sample run through the extraction protocol simultaneously with the actual samples and a negative PCR control in which the template DNA was replaced with sterile H_2O . These remained free of PCR amplicons.

A total of 20 μ l of each of the three amplicons for each sample was combined and purified using an AmPure SPRI (AgenCourt Bioscience, Beverly, MA, USA) magnetic PCR clean-up system. This clean-up system was selected because it discriminates against fragments of less than 100 bp in size and effectively eliminates dimers of the fusion primer constructs that exceed 40 bp in size. The clean PCR products were again quantified with the ND1000 spectrometer, 50 ng of clean DNA per sample was combined for sequencing and this pool was adjusted to 10 ng μ l⁻¹. The pooled products were sequenced in two 1/16th regions of a sequencing reaction of a GS-FLX sequencer (454 Life Sciences, Branford, CT, USA) at the Interdisciplinary Center for Biotechnology Research at the University of Florida. We expected an average sequencing depth of 370 reads for each of the 54 samples, assuming a 10 000 sequence yield for each of the 1/16th regions. Raw data are available from the Genome Short Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) under accession number 34493.

Bioinformatics and operational taxonomic unit (OTU) designation

The generated sequences were searched for the occurrence of the DNA tag immediately preceding the ITS2 primer sequence. When found, the DNA tag was removed and the

corresponding sample designation was incorporated into the fasta sequence identifier. For quality control, sequences were removed if they contained no valid primer sequence or DNA tag, contained ambiguous bases or were shorter than our length threshold (200 bp). The sequences that met these quality control criteria were merged across the two sequencing regions to generate a single input for subsequent analyses. The sequences were aligned with CAP3 (Huang & Madan, 1999) and assigned to OTUs at 69–99% similarity at 2% intervals using a minimum overlap of 100 bp. All singletons as well as example sequences for each OTU at 99% similarity are available at GenBank (accession numbers FJ756950–FJ763141). The data were parsed by sample to calculate the OTU frequencies for each sample. From this output, SAS (SAS Institute Inc., Cary, NC, USA) was used to calculate richness and diversity indices.

Diversity indices

The overall OTU richness (S) was calculated by summing the number of OTUs, including singletons, within each sample. Simpson's dominance ($D = \sum p_i^2$), Simpson's diversity ($1 - \sum p_i^2$) and Shannon's diversity ($H' = -\sum p_i[\log_e(p_i)]$) were calculated for each sample, where p_i is the frequency of occurrence of each OTU. Evenness was calculated as the ratio of Shannon's diversity and richness (H'/S). A final index of diversity, Fisher's alpha log-series (Fisher *et al.*, 1943), was calculated by iterating the equation $S/N = [(1-x)/x] [-\log_e(1-x)]$, where S is richness and N is the total number of sequences within the sample. Once x was solved, the diversity index alpha (α) was calculated as $N(1-x)/x$. To explore organismal coverage among the focal trees and land-use types, species' accumulation (rarefaction) curves and extrapolative richness estimators were generated using EstimateS (version 8; Colwell, 2006).

OTU frequency

OTU frequencies were analyzed for OTUs assigned at 95% similarity level. This sequence similarity was chosen based on taxon accumulation as a function of sequence similarity (Fig. 2a,e): although the OTU richness is relatively stable up to 95% sequence identity it assumes near-exponential growth at thresholds of more than 95%. It is also noteworthy that our sequences only include ITS1 and contain little conserved subunit sequences (3' end of 18S or 5' end of 5.8S). We used two different strategies to identify taxa that may differ across the land-use types. First, all nonsingleton OTUs were analyzed using JMP 7.0.1 (SAS Institute Inc.). Second, a representative of each nonsingleton OTU was assigned to a genus based on BLAST matches (Zhang *et al.*, 2000). OTU frequencies were summed across each experimental unit and genus, and analyzed for effects of land use on this level. We acknowledge that the inaccuracies in database annotations

(Nilsson *et al.*, 2006; Arnold *et al.*, 2007), as well as the separate accessioning of teleomorph and anamorph genera, present potential problems. Nonetheless, we chose this approach because unambiguous sequence alignment across ITS1 is difficult. Furthermore, the genus-level assignment provides further information beyond the simple OTU level and is preferable to the family level, as many culturable anamorph taxa have not been appropriately assigned to higher taxonomic levels. As a result, we argue that, on this level, we could better summarize the results and reduce the data volume. To account for the effect of poor matches in the BLAST analyses, we examined whether or not our analyses remained consistent after removal of those data points whose assignment was based on short overlap (< 80% coverage) or low sequence similarity (< 90% similarity) among the query and match sequences.

Statistical analyses

The three leaves from the three focal trees at each site do not represent truly independent samples. Accordingly, we averaged the measured dependent variables for each tree ($n = 18$). To account for variability among sites, we analyzed the responses using a nested analysis of variance (ANOVA) model with independent variables 'land use' and 'site(land use)' in JMP 7.0.1 (SAS Institute). To estimate the contribution of each explanatory variable, the variance components were partitioned and the proportion of the total variance represented by each explanatory variable was calculated in JMP 7.0.1. There is a further concern as to whether or not the clustered distribution of trees within a stand truly constitutes a tree as an experimental unit. To test the consistency of our conclusions, we also used one-way ANOVA with the 'site' ($n = 6$) as an experimental unit and 'land use' as the independent variable.

To determine the effects of sequence similarity threshold on diversity and richness estimators, the data were analyzed in two ways. First, in a universal analysis that incorporated the similarity threshold ('similarity') as a categorical class variable, we examined the interaction terms between 'similarity' and 'land use' or 'site(land use)'. It should be noted that this strategy exaggerates the sampling effort 16-fold and should not be used for main effect analyses. The nonsignificant interaction terms were interpreted as an indication of additive main effects. Second, as the first analysis clearly presents some problems, the data were also analyzed separately for each level of 'similarity'. The conclusions about the 'land use' effects were particularly examined to confirm whether or not these conclusions remained invariable across the levels of 'similarity'.

To examine differences in the phyllosphere fungal communities among the land-use types, the frequencies of the 125 observed genera were analyzed in PC-ORD (v. 4.1, McCune & Mefford, 1999). The genus-level frequencies were selected because the analyses of individual OTUs saturated the ordination space. Pairwise community distances on a tree level

were estimated using the Sørensen (Bray–Curtis) index and analyzed using nonmetric multidimensional scaling (NMS; Mather, 1976) multivariate analysis. The optimal number of dimensions (k) was selected based on the Monte Carlo test of significance at each level of dimensionality comparing 40 runs with empirical data against 50 randomized runs with a step-down in dimensionality from six to one and a random seed starting value. The $k \geq 2$ dimensional solutions yielded similar results and produced solutions with stress values smaller than those in randomized runs ($P = 0.0196$). Accordingly, the two-dimensional solutions were selected and the data were re-ordinated with a $k = 2$ configuration.

Results

After quality control and removal of 2707 reads that did not meet our minimum requirements, we retained 9168 high-quality sequences for the trees within the city of Manhattan and 8852 for the trees outside the city (average sequence length, 265 bp). The total number of acquired sequences per sample was invariable among the land uses (ANOVA: $F_{1,16} = 0.0028$ $P = 0.9580$) and sites within land use (ANOVA: $F_{4,18} = 0.9953$ $P = 0.4319$), indicating successful and accurate pooling of the templates. At 95% similarity level, these 18 020 sequences represented high richness: 329 singleton OTUs and 360 nonsingleton OTUs (i.e. 689 OTUs in all).

Despite their small size (3.14 cm² per leaf), the samples hosted high species' richness and diversity: rarefaction analyses of the OTUs assigned at 95% similarity (Fig. 1) indicated that none of the samples were saturated for organismal coverage, even in the trees with deepest sequencing effort (> 1700 reads per tree). Extrapolative richness estimators (Chao1, first-order jackknife) corroborate and provide, on average, values two-fold greater than the observed richness (Table S3, see Supporting Information). Accordingly, far greater sequencing effort would be necessary to reach a plateau in phyllosphere fungal richness.

The oak trees in rural sites hosted greater OTU richness (S) and diversity (Fisher's α) in their phyllosphere fungal communities than did the trees within the city (Figs 1, 2; Table 2) regardless of the choice of experimental unit (Fig. 2). Shannon's (H) and Simpson's ($1 - D$) diversity indices were less consistent and indicated similar differences only when a tree was considered as an experimental unit (compare Fig. 2c,d,g,h) and at sequence similarities of less than 87% (Fig. 2c). Evenness (H'/S) was invariable across the land-use types. The land use effect was strong, as indicated by our evaluation of variance partitioning among the factors (variance component analysis). To exemplify, in a nested model that also accounted for individual trees within a site, the within-tree variance constituted 49.2% of the total for richness and 14.8% of the total for diversity (Fisher's α), followed by the land-use effect (37.1% for richness and 57.1% diversity), sites within the land use

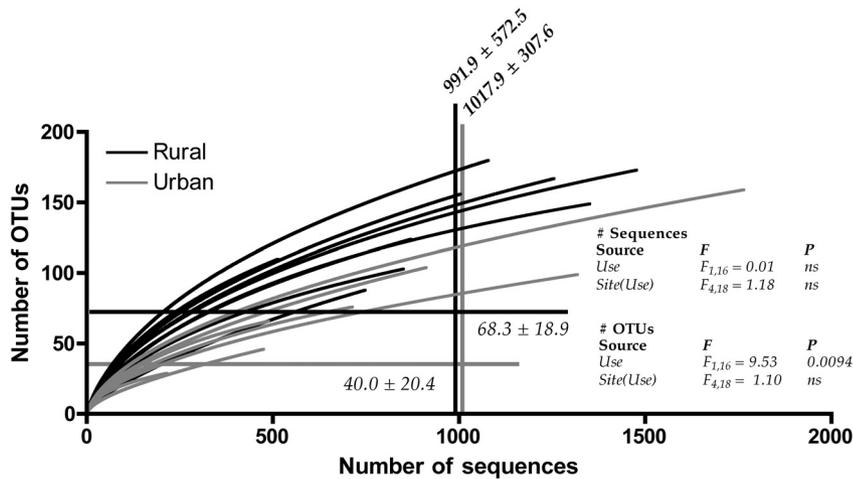


Fig. 1 Rarefaction curves for each of the 18 trees analyzed (nine *Quercus macrocarpa* trees in two contrasting land uses): rural, black line; urban, grey line. Operational taxonomic units (OTUs) were assigned at 95% similarity for these analyses. The two included analysis of variance (ANOVA) tables show the analyses at this threshold when a tree is considered as an experimental unit. Vertical lines indicate the number of sequences per sample [mean \pm one standard deviation (1SD)]; horizontal lines indicate the number of OTUs per sample (mean \pm 1SD).

Table 2 Analysis of variance (ANOVA) tables for the phyllosphere fungal responses to land use

Response	Model: $Y = \text{land use} + \text{site(used)}$		Model: $Y = \text{use}$
	Land use $F_{1,16}^P$	Site(used) $F_{4,18}^P$	Land use $F_{1,4}^P$
Number of sequences	0.01 ^{0.9068}	0.97 ^{0.4598}	0.01 ^{0.9092}
OTU richness (<i>S</i>)	9.53 ^{0.0094**}	1.10 ^{0.4013}	8.68 ^{0.0422*}
Singletons	3.39 ^{0.0905}	1.32 ^{0.3187}	2.57 ^{0.1840}
Diversity indices			
Fisher's α	18.61 ^{0.0010**}	1.33 ^{0.3145}	13.99 ^{0.0201*}
Simpson's (<i>D</i>)	2.89 ^{0.1150}	2.84 ^{0.0717}	3.41 ^{0.1385}
Shannon's (<i>H'</i>)	10.38 ^{0.0073**}	2.75 ^{0.0781}	3.78 ^{0.1239}
Evenness (<i>H'/S</i>)	1.30 ^{0.2768}	1.34 ^{0.3123}	0.97 ^{0.3802}

A nested model that accounts for site effects and in which a tree is considered as an experimental unit and another model in which site is considered as an experimental unit are presented. For these analyses, the operational taxonomic units (OTUs) were assigned at 95% similarity level in CAP3. Land-use effects for both models across multiple sequence similarity thresholds are presented in Fig. 2. Responses that are significant at $\alpha < 0.05$ are given in bold for emphasis. *, $0.01 < P \leq 0.05$; **, $0.001 < P \leq 0.01$; ***, $P \leq 0.001$.

(7.7% for richness and 3.3% diversity) or among trees within a site (6.0% for richness and 24.8% diversity). Accordingly, the land-use effect accounted for more than four times the variance represented by the site effects for richness and diversity.

To account for the potential effects of OTU threshold selection, we assigned the OTUs at 16 different levels of sequence similarity (69–99% at 2% intervals; Fig. 2). Examination of the interaction terms (Table S4, see Supporting Information) showed that the differences in richness, diversity indices and evenness were additive and consistent among the land-use types and sites within land-use type across the classes of sequence similarity. Similarly, the conclusions on the differences between the two land uses were consistent at most levels of sequence similarity when analyzed independently at each level (Fig. 2). One should interpret the sliding scale of similarity in a phylogenetic or taxonomic context. The consistent differences at lower levels of sequence similarity suggest

that the observed patterns are indicative of greater richness and diversity at higher taxonomic ranks, i.e. not only on species, but also on genus and/or family levels.

The acquired sequences represented largely two fungal phyla: Ascomycota (284 OTUs; 93.4% of the sequences) were overwhelmingly most frequent, followed by Basidiomycota (73 OTUs; 6.5%). One OTU representing the basal fungal lineages (subphylum Mucoromycotina) was also detected (one OTU; 0.02%). Phylum level placement could not be inferred from BLAST analyses for five sequences (0.03% of all sequences), and 18 OTUs (24.3% of all sequences) could only be assigned as ascomycetes. These data indicate that, for these sequences, no well-accessioned, closely related fungi were available for comparisons in the databases.

To provide an additional, conservative taxon placement by BLAST analysis, we filtered our data for coverage ($\geq 80\%$ of the sequence length) and similarity ($\geq 90\%$ sequence

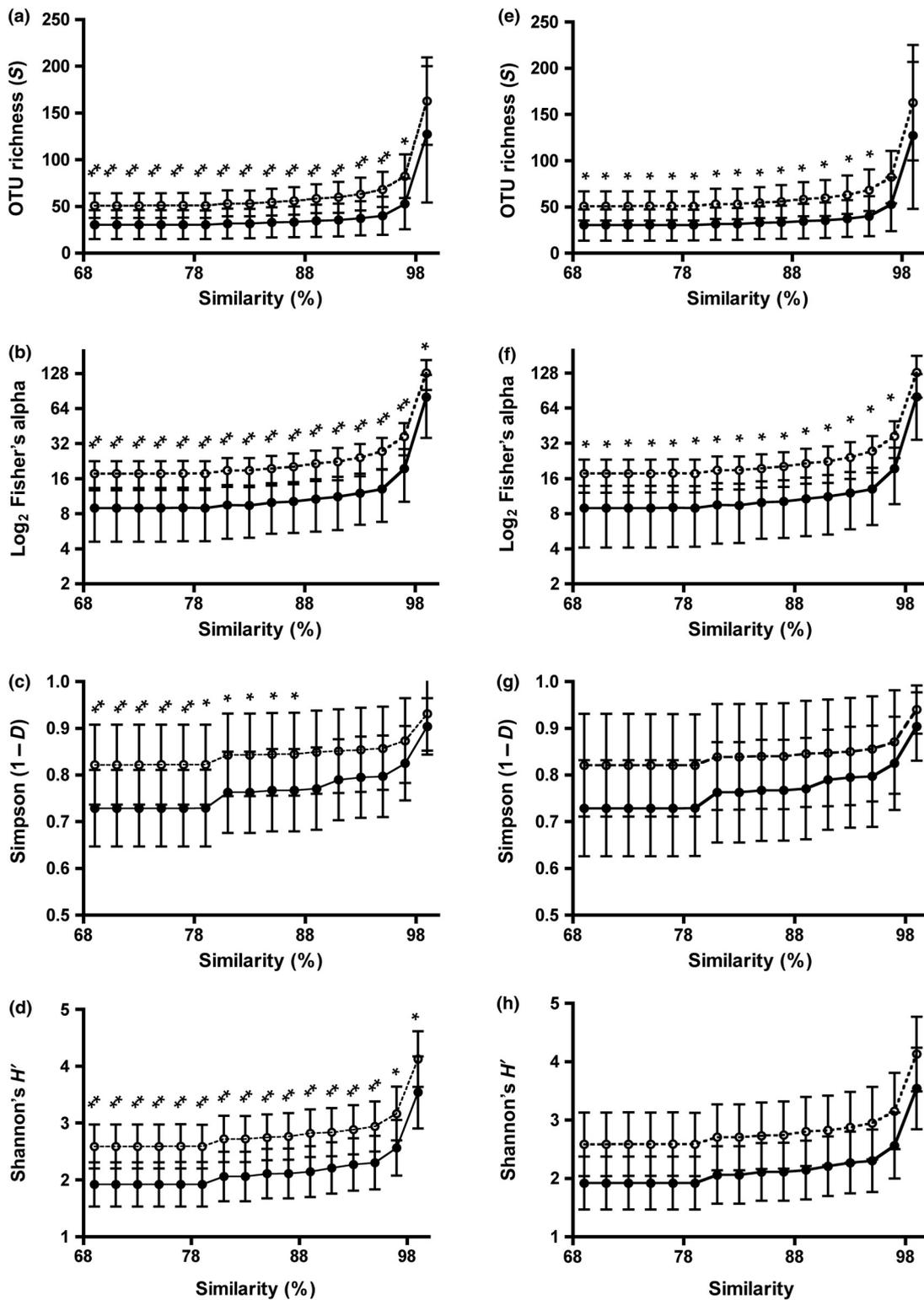


Fig. 2 *Quercus macrocarpa* phyllosphere fungal communities within (urban) and outside (rural) the city limits of Manhattan, KS, USA (open circles, nonurban; filled circles, urban). For figures a–d, a tree was considered an experimental unit ($n = 18$); for figures e–h, a site ($n = 6$). (a, e) Operational taxonomic unit (OTU) richness (S); (b, f) Fisher's α diversity index; (c, g) Simpson's diversity index ($1 - D$); (d, h) Shannon's diversity index (H'). Error bars indicate \pm one standard deviation (1SD). Table S3 (see Supporting Information) shows the overall analysis of variance (ANOVA) tables across the 16 OTU assignment levels, indicating the lack of interaction among similarity and land use. The asterisks indicate the level of significance at each of the similarity thresholds for OTU assignment; nonsignificant comparisons are not shown; * $0.01 \leq P \leq 0.05$; ** $0.001 \leq P < 0.01$.

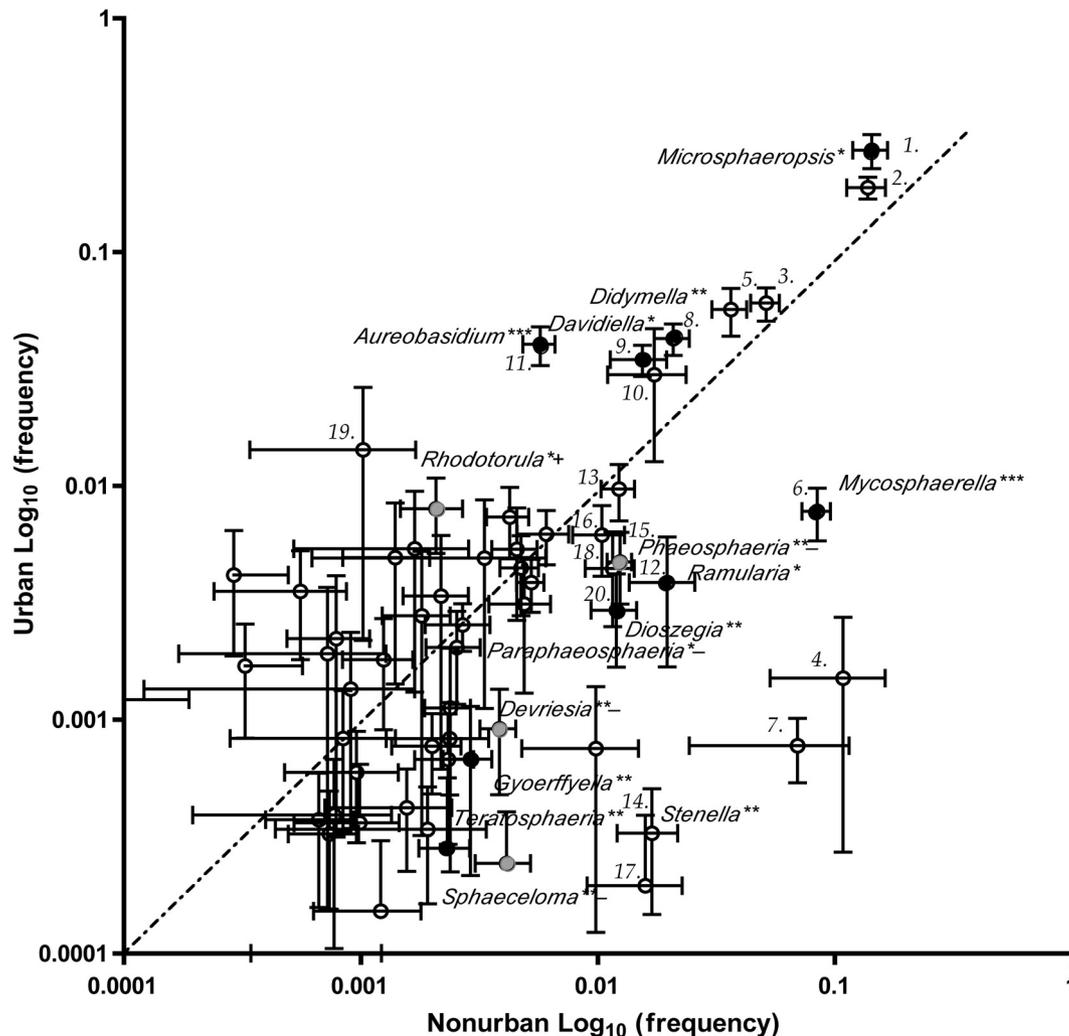


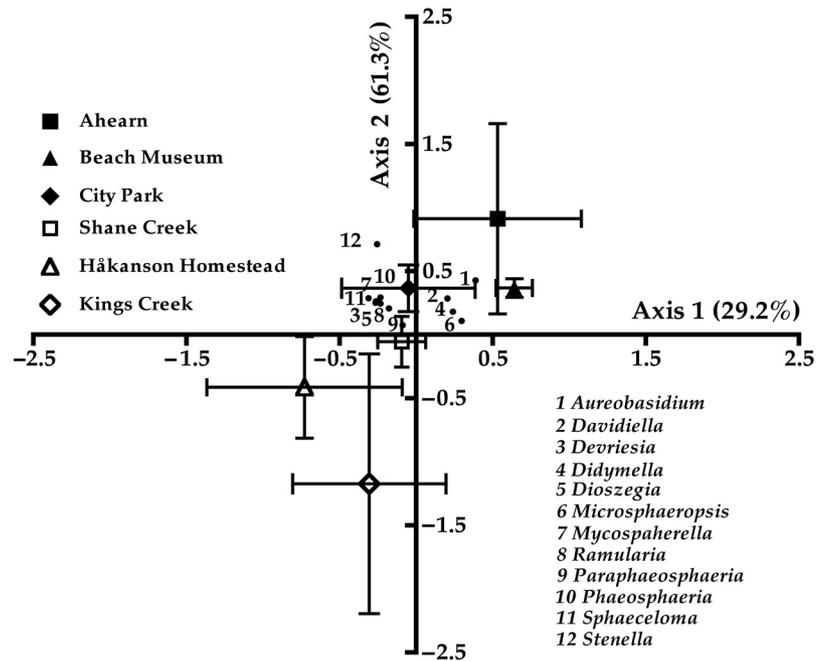
Fig. 3 Occurrence [mean \pm one standard deviation (1SD)] of the 63 detected genera outside (x-axis) and within (y-axis) the city limits of Manhattan, KS, USA. The 63 genera were selected by omitting taxa that occurred in four or fewer sites. Both axes are \log_{10} -scaled. The broken line indicates equal frequency in both land-use types; taxa above the line are more frequent within the city limits, taxa below the line are more frequent outside. Filled black symbols highlight taxa that occur at different frequencies at $P < 0.05$; grey symbols indicate those affected by filtering by query coverage and similarity in BLAST analyses (see text). Numbers identify the 20 most frequent genera. 1, *Microsphaeropsis*; 2, *Alternaria*; 3, *Epicoccum*; 4, *Erysiphe*; 5, *Cladosporium*; 6, *Mycosphaerella*; 7, *Taphrina*; 8, *Didymella*; 9, *Davidiella*; 10, *Neofabraea*; 11, *Aureobasidium*; 12, *Ramularia*; 13, *Cryptococcus*; 14, *Stenella*; 15, *Phaeosphaeria*; 16, *Filobasidium*; 17, *Lalaria*; 18, *Sporobolomyces*; 19, *Exophiala*; 20, *Dioszegia*. * $0.01 \leq P < 0.05$; ** $0.001 \leq P < 0.01$; *** $P < 0.001$.

identity). After filtering, 160 OTUs (11.3% of the sequences) were considered to be uncertain in their placement based on their low coverage and/or low similarity to accessioned sequences. The remaining OTUs were distributed across 28 orders, 33 families and 81 genera. Among the most commonly encountered orders were Pleosporales (48 OTUs; 26.7% of all sequences), Capnodiales (44 OTUs; 12.8%), Erysiphales (three OTUs; 7.9%), Tremellales (15 OTUs; 2.2%) and Taphrinales (nine OTUs; 4.2%). On the family level, Pleosporaceae (25 OTUs; 17.5%), Leptosphaeriaceae (13 OTUs; 9.0%), Erysiphaceae (three OTUs; 7.9%), Davidiellaceae (nine OTUs; 6.9%), Taphrinaceae (nine OTUs;

4.2%) and Mycosphaerellaceae (30 OTUs; 5.8%) were the most common.

At genus level, filtering reduced the number of putative genera from 125 to 81. The most common genera (see Fig. 3 for top 20) were *Microsphaeropsis* (two OTUs; 23.6%), *Alternaria* (six OTUs; 16.1%), *Epicoccum* (one OTU; 6.0%), *Erysiphe* (two OTUs; 5.9%), *Cladosporium* (five OTUs; 4.5%), *Mycosphaerella* (24 OTUs; 4.4%) and *Taphrina* (eight OTUs; 3.4%). Filtering by BLAST coverage and similarity did not change the ranking of the most common taxa, but often decreased the number of OTUs within a taxon, as well as the sequences placed within that taxon. For example,

Fig. 4 Nonparametric multidimensional scaling (NMS) ordination of the *Quercus macrocarpa* phyllosphere fungus communities. Plotted are the scores [mean \pm one standard deviation (1SD)] for the sites. Open symbols, sites outside the city limits of Manhattan, KS, USA; filled symbols, sites within the city limits of Manhattan, KS, USA. The percentages following the axes indicate the proportion of variation represented by that axis. Small dots identify genera that were determined to be responsive in the analysis of variance (ANOVA), and the numbers refer to the list of genera in the inset.



within Pleosporales, filtering reduced the number of OTUs from 61 (27.7% of all sequences) to 48 (26.7% of all sequences). By contrast, the number of OTUs within Erysiphales and Taphrinales remained unchanged. In the case of Erysiphales, the number of reads remained the same, whereas filtering reduced the occurrence of Taphrinales from 4.8 to 4.5%. Taken together, the filtering mainly removed OTUs that were relatively low in frequency.

A total of 41 OTUs (11.4% of all nonsingletons) occurred at different frequencies (ANOVA: $P < 0.05$) among the land-use types. These OTUs largely represented the most commonly occurring genera, and we present only the genus-level analyses here [but see Table S2 (Supporting Information) for responsive OTUs]. Analyses on the genus level (Fig. 3) showed that *Devriesia*; *Mycosphaerella*, its anamorph *Ramularia*, and *Stenella* (Capnodiales); *Dioszegia* (Tremellales); *Paraphaeosphaeria* and *Phaeosphaeria* (Pleosporales); and *Sphaeceloma* (Myriangiales) were more frequent in the nonurban environments. By contrast, *Aureobasidium* (mitosporic Chaetothryiomycete), *Davidiella* (Capnodiales), *Didymella* (mitosporic Dothideomycete) and *Microsphaeropsis* (unknown) occurred more frequently in the urban environments. We analyzed both the filtered and nonfiltered datasets at this level of taxonomic resolution. Although the number of considered genera was reduced (above), the conclusions on generic frequencies were generally congruent among these datasets (Fig. 3). Filtering added *Rhodotorula* to the genera that were more frequent in the urban environments and removed *Devriesia*, *Paraphaeosphaeria*, *Phaeosphaeria* and *Sphaeceloma* from those that occurred at different frequencies among the land-use types.

We also compared the overall community composition among the urban and nonurban environments using NMS. The two first ordination axes represented a substantial 90.5% of the variation and separated the urban and nonurban communities (Fig. 4). The genera that differed in frequency among the two land-use types were placed near the origin of the ordination space, indicating that the differences were driven by the overall composition rather than by the few taxa that differed in their frequencies.

Discussion

To our knowledge, we have described the first successful attempt to combine DNA tagging and direct pyrosequencing of fungal communities to test hypotheses on phyllosphere fungal richness and diversity. Our deeply partitioned sequencing using two 1/16th regions of a GS-FLX aimed not to maximize the number of reads per sample, but rather to cost-effectively optimize the number of reads (on average, ~1000 reads per tree) and number of samples that could be analyzed simultaneously (54 samples representing 18 trees). We consciously compromised between maximal sequencing depth and statistical power to test our primary hypotheses. We emphasize that one must make these decisions with the primary goals of a project in mind: although deeper sequencing may be necessary to improve the accuracy of the richness estimators, a greater number of samples will improve the ability to make statistical inferences. As evidenced by this study and those targeting bacterial 16S in environmental studies (for example, Sogin *et al.*, 2006; Roesch *et al.*, 2007;

Huse *et al.* 2008), these attempts often require far more extensive sequencing than anticipated.

Our analysis of phyllosphere fungal communities indicated that temperate deciduous oaks host a high species' richness and diversity. Several authors have suggested that leaf-associated communities are diverse (Fröhlich & Hyde, 1999; Arnold *et al.*, 2000; Santamaria & Bayman, 2005), partly because of the highly localized colonization of the foliar tissues (Lodge *et al.*, 1996). Our results indicate that, although we detected nearly 700 phyllosphere fungi, the potential richness is probably more than two-fold higher. Extrapolative measures provided values nearly twice the observed richness, but tend to be negatively biased if the true richness substantially exceeds the observed richness, or when the sampling does not provide adequate representation (Palmer, 1990; Baltanas, 1992).

For comparison with our results, Arnold *et al.* (2007) analyzed 90 surface-sterilized *Pinus taeda* needles from Duke Forest in North Carolina. They cultured 439 isolates, which represented 37 ITS OTUs at 95% similarity when 150 isolates were sampled for sequencing. In a study on a tropical island in Panama, Arnold *et al.* (2000) pure cultured 418 morphospecies from 83 leaves of two host species, which were concluded to represent a total of 347 genetically distinct taxa, or approximately 200 genetically distinct species per host species. More than one-half of those taxa were represented by a single pure-cultured isolate, leading to an argument for hyperdiverse foliar endophyte communities in tropical forests. Our study and those of Arnold and collaborators are not directly comparable because of different targets, methods and environments (see Santamaria & Bayman, 2005; Arnold *et al.*, 2007). However, it is tempting to draw parallels. In our study, as in that of Arnold *et al.* (2000), approximately one-half of the detected fungi were singletons. Our sampling concentrated on a single temperate host with a limited sampling of only 54 leaves, at 95% sequence similarity, but contained both epiphytes and endophytes. As a result, including both epiphytes and endophytes as well as deeper sampling, our estimated number of molecular OTUs is more than three times that estimated by Arnold *et al.* (2000) for either of their two tropical hosts. Direct comparisons using similar methods would be interesting to elucidate the patterns in taxon richness among temperate and tropical environments. Regardless, the bottom line remains: foliar habitats in tropical and temperate environments host a tremendous fungal diversity: with our sampling of up to > 1700 sequences per tree, we did not capture the existing phyllosphere fungal richness.

The observed fungal richness and diversity varied dramatically among the land uses. Furthermore, the frequencies of many individual OTUs (> 10%) and BLAST-inferred genera also differed. The land-use effects were surprisingly strong considering the small size of the urban development with *c.* 50 000 permanent residents. We expect these effects to be even more pronounced in a more substantial urban

development. It is also of note that the effects of land use were visible even when the sites (i.e. not individual trees) were considered as the experimental unit. The underlying reasons for the observed differences remain uncertain and are subject to ongoing investigations. In brief, there are a number of possible drivers, including differences in host population structure or genetic variation (Pan *et al.*, 2008), stand isolation and size (Helander *et al.*, 2007), edaphic properties and differences in nutrient availability within the sampled stands (Pouyat & McDonnell, 1991; Craul, 1992; Kaye *et al.*, 2006), as well as differences in management choices among the stands that alter the stressors to trees (Shochat *et al.*, 2004) or reduce residual plant debris that may provide a source of fungal propagules (Sutton, 1992; Herre *et al.*, 2007).

To reduce the data volume and to provide a taxonomic framework, we used BLAST analyses to assign OTUs to genera and performed further analyses at this level of resolution. The choice of taxonomic level above that of a species also decreases the probability of erroneous taxon assignments stemming from incongruence among analyzed fragments (Nilsson *et al.*, 2009). In our analyses, nonlichenized, filamentous ascomycetes dominated the observed fungal communities (see also Arnold, 2008) and represented anamorphs and teleomorphs that are common in air samples (for example, *Alternaria*, *Aureobasidium*, *Cladosporium*, *Davidiella*, *Didymella*, *Epicoccum*) and in living or dead plant materials (*Alternaria*, *Aureobasidium*, *Exophiala*, *Microsphaeropsis*). Functionally, the detected fungi represented pathogens or abundant putative endophytes (*Alternaria*, *Aureobasidium*, *Davidiella*, *Didymella*, *Microsphaeropsis*, *Mycosphaerella* and its anamorph *Ramularia*, *Neofabrea*, *Phaeosphaeria*, *Stenella*, *Taphrina* and its anamorph *Lalaria*), powdery mildews (*Erysiphe* and its anamorph *Oidium*) and common foliar fungi that may utilize leaf exudates (the basidiomycetous yeasts *Sporobolomyces* and *Dioszegia* or *Taphrina* and its anamorph *Lalaria*).

Although we argue that taxon assignment that accounts for coverage and similarity provides tangible information, it also presents problems. Reliance on databases leads to inherent accessioning bias – organisms of economic importance are well sampled and often provide excellent annotations, whereas organisms with unknown economic importance remain poorly identified and may often be erroneously annotated. The inadequate and erroneous annotations (Nilsson *et al.*, 2006), as well as changes in taxonomic treatise and the parallel taxonomies facilitating the teleomorph and anamorph taxa, present further problems that may vary across morphological taxa depending on how well the morphological species concepts are understood (Arnold *et al.*, 2007). In addition, it remains unclear how variation in BLAST similarities and coverage should be treated. We chose to remove the sequences with less than 80% coverage and 90% similarity. Obviously, taxon assignment using such filters may alter the abundance estimators, but does provide information beyond an unidentified OTU.

The use of short reads, similarly, presents potential problems. Liu *et al.* (2008) and Nilsson *et al.* (2009) examined how well the short pyrosequenced reads perform in taxon assignment. Although they concluded that short reads provide adequate information for taxon assignment, the assignment was sensitive to the region targeted. Problematically, use of ITS1 or ITS2 regions resulted in incongruent species-level taxon assignments in 40% of cases (Nilsson *et al.*, 2009). Even with the problems identified, the number of accessions, organismal coverage and common use in barcoding strongly favor ITS as a target. Although the short and variable ITS seemed to serve well for taxon assignment in our study, the inference from these data would have been improved if tools, such as those available for 16S reads in the Ribosomal Database Project (<http://rdp.cme.msu.edu/>) or GreenGenes (<http://greengenes.lbl.gov/>), were available. However, unambiguous alignment of the ITS regions is difficult, especially in the absence of conserved regions. To somewhat account for these difficulties, we performed OTU binning at multiple similarity thresholds and chose 95% similarity, as richness estimators tended to remain relatively stable and our conclusions largely invariable below this level. Such strategies may be necessary until better and more broadly utilizable tools become available.

To conclude, our pyrosequencing results strongly indicate that the temperate trees host extremely diverse fungal communities in their phyllosphere. Most of the detected fungi represented genera that either are common in aerial samples or frequently occupy the phyllosphere, and represent a number of ecological roles, including pathogens, latent saprobes and endophytes. The functional roles and life stages (spores, germinating spores, established hyphae, etc.) of the detected fungi are almost impossible to determine in a molecular assay such as ours. However, many of the most abundant genera represented culturable genera that have been previously observed in assays targeting endophytes. In those studies, similar to that described here, the ecological roles of the acquired fungal isolates have remained unconfirmed and probably represent taxa that may or may not form endophyte symbioses.

Acknowledgements

This research was funded by Kansas State University, Division of Biology and Ecological Genomics Institute. The authors are indebted to Dr Charles L. Kramer for his assistance in field sampling as well as a critical review of an earlier manuscript. J. David Mattox from the City of Manhattan assisted in stand selections, tree identifications and sampling. Konza Prairie Long-Term Ecological Research provided access to nonurban stands for comparisons. Regina Shaw, Interdisciplinary Center for Biotechnology Research at the University of Florida, performed 454 sequencing at the University of Florida Genomics Core Facility.

References

- Arnold AE. 2007. Understanding the diversity of foliar fungal endophytes: progress, challenges and frontiers. *Fungal Biology Reviews* 21: 51–66.
- Arnold AE. 2008. Endophytic fungi: hidden components of tropical community ecology. In: Schnitzer S, Carson W, eds. *Tropical forest community ecology*. Chichester, UK: John Wiley & Sons, 254–271.
- Arnold AE, Henk DA, Eells RL, Lutzoni F, Vilgalys R. 2007. Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia* 99: 185–206.
- Arnold AE, Lutzoni F. 2007. Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* 88: 541–549.
- Arnold AE, Maynard Z, Gilbert GS, Coley PD, Kursar TA. 2000. Are tropical endophytes hyperdiverse? *Ecology Letters* 3: 267–274.
- Baltanas A. 1992. On the use of some methods for the estimation of species richness. *Oikos* 65: 484–492.
- Blair RB. 1999. Birds and butterflies along an urban gradient: surrogate taxa for assessing biodiversity? *Ecological Applications* 9: 164–170.
- Blair RB. 2004. The effects of urban sprawl on birds at multiple levels of biological organization. *Ecology and Society* 9: 2.
- Carroll GC, Muller EM, Sutton BC. 1977. Preliminary studies on the incidence of needle endophytes in some European conifers. *Sydowia* 29: 87–103.
- Cincotta RP, Engelman R. 2000. Biodiversity and population growth. *Issues in Science and Technology* 16: 80–81.
- Colwell RK. 2006. *EstimateS: Statistical estimation of species richness and shared species from samples*. Version 8. <http://purl.oclc.org/estimates>
- Craul PJ. 1992. *Urban soil in landscape design*. New York, NY, USA: John Wiley & Sons.
- Fisher RA, Corbet AS, Williams CB. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 12: 42–58.
- Fröhlich J, Hyde KD. 1999. Biodiversity of palm fungi in the tropics: are global fungal diversity estimates realistic? *Biodiversity and Conservation* 8: 977–1004.
- Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2: 113–118.
- Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods* 5: 235–237.
- Helander M, Ahlholm J, Sieber TN, Hinneri S, Saikkonen K. 2007. Fragmented environment affects birch leaf endophytes. *New Phytologist* 175: 547–553.
- Herre EA, Mejia LC, Kyllö DA, Rojas E, Maynard Z, Butler A, van Bael SA. 2007. Ecological implications of anti-pathogen effects of tropical fungal endophytes and mycorrhizae. *Ecology* 88: 550–558.
- Huang XQ, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Research* 9: 868–877.
- Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics* 4: e1000255.
- Kaye JP, Groffman PM, Grimm NB, Baker LA, Pouyat RV. 2006. A distinct urban biogeochemistry? *Trends in Ecology and Evolution* 21: 192–199.
- Kühn I, Brandl R, Klotz S. 2004. The flora of German cities is naturally species rich. *Evolutionary Ecology Research* 6: 749–764.
- Lambais MR, Crowley DE, Cury JC, Bull RC, Rodrigues RR. 2006. Bacterial diversity in tree canopies of the Atlantic forest. *Science* 312: 1917.
- Liu ZZ, DeSantis TZ, Andersen GL, Knight R. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research* 36: e120.

- Lodge DJ, Fisher PJ, Sutton BC. 1996. Endophytic fungi of *Manilkara bidentata* leaves in Puerto Rico. *Mycologia* **88**: 733–738.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Mather PM. 1976. *Computational methods of multivariate analysis in physical geography*. London, UK: J. Wiley & Sons.
- McCune B, Mefford MJ. 1999. *PC-ORD multivariate analysis of ecological data, version 4*. Glendale Beach, OR, USA: MjM Software.
- McKinney ML. 2002. Urbanization, biodiversity, and conservation. *Bioscience* **52**: 883–890.
- McKinney ML, Lockwood JL. 1999. Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends in Ecology and Evolution* **14**: 450–453.
- Meyer M, Stenzel U, Hofreiter M. 2008. Parallel tagged sequencing on the 454 platform. *Nature Protocols* **3**: 267–278.
- Nilsson RH, Ryberg M, Abarenkov K, Sjökvist E, Kristiansson E. 2009. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiology Letters* **296**: 97–101.
- Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Kõljalg U. 2006. Taxonomic reliability of DNA sequences in public databases: a fungal perspective. *PLoS ONE* **1**: e59.
- Palmer MW. 1990. The estimation of species richness by extrapolation. *Ecology* **71**: 1195–1198.
- Pan JJ, Baumgarten AM, May G. 2008. Effects of host plant environment and *Ustilago maydis* infection on the fungal endophyte community of maize (*Zea mays*). *New Phytologist* **178**: 147–156.
- Pouyat RV, McDonnell MJ. 1991. Heavy metal accumulations in forest soils along an urban–rural gradient in southeastern New York, USA. *Water Air and Soil Pollution* **57–8**: 797–807.
- Rodrigues RJ, White JF, Arnold AE, Redman RS. 2009. Fungal endophytes: diversity and functional roles. *New Phytologist* **182**: 314–330.
- Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, Daroub SH, Camargo FAO, Farmerie WG, Triplett EW. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME Journal* **1**: 283–290.
- Saikkonen K, Faeth SH, Helander M, Sullivan TJ. 1998. Fungal endophytes: a continuum of interactions with host plants. *Annual Review of Ecology and Systematics* **29**: 319–343.
- Santamaria J, Bayman P. 2005. Fungal epiphytes and endophytes of coffee leaves (*Coffea arabica*). *Microbial Ecology* **50**: 1–8.
- Shochat E, Lerman SB, Katti M, Lewis DB. 2004. Linking optimal foraging behavior to bird community structure in an urban-desert landscape: field experiments with artificial food patches. *American Naturalist* **164**: 232–243.
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored ‘rare biosphere’. *Proceedings of the National Academy of Sciences, USA* **103**: 12 115–12 120.
- Sutton JC. 1992. Epidemiology of wheat head blight and maize ear rot caused by *Fusarium graminearum*. *Canadian Journal of Plant Pathology* **4**: 195–209.
- Wania A, Kuhn I, Klotz S. 2006. Plant richness patterns in agricultural and urban landscapes in Central Germany – spatial gradients of species richness. *Landscape and Urban Planning* **75**: 97–110.
- White TJ, Bruns TD, Lee SB, Taylor JW. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols: a guide to methods and applications*. New York, NY, USA: Academic Press, 315–322.
- Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7**: 203–214.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 List of the internal transcribed spacer (ITS) primers that combine fungus specific (ITS1F; Gardes & Bruns, 1993) and universal (ITS2; White *et al.*, 1990) primers and the 454-sequencing endemic A- and B-primers (Margulies *et al.*, 2005) plus a 5-bp DNA tag for post-sequencing sample identification

Table S2 Operational taxonomic unit (OTU) assignments before and after filtering for coverage and similarity

Table S3 Operational taxonomic unit (OTU) richness estimators, observed number of OTUs, number of singletons, proportion of singletons, number of doubletons and two extrapolative estimators (Chao 1 and first-order jackknife)

Table S4 Analysis of variance (ANOVA) tables for the phyllosphere fungal responses to land use

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.